



ΠΑΝΕΠΙΣΤΗΜΙΟ ΘΕΣΣΑΛΙΑΣ

**ΣΧΟΛΗ ΘΕΤΙΚΩΝ ΕΠΙΣΤΗΜΩΝ
ΤΜΗΜΑ ΠΛΗΡΟΦΟΡΙΚΗΣ ΜΕ ΕΦΑΡΜΟΓΕΣ ΣΤΗ
ΒΙΟΪΑΤΡΙΚΗ**

**Συσχέτιση miRNAs με καρκινικούς φαινοτύπους μέσω ανάλυσης
δεδομένων από CRISPR-Cas9 Knockout screens**

Βασιλική Πελεκάνου

Διπλωματική εργασία

**Επιβλέπουσα Καθηγήτρια:
Καθ. Άρτεμις Χατζηγεωργίου**

Λαμία, 2023



ΠΑΝΕΠΙΣΤΗΜΙΟ ΘΕΣΣΑΛΙΑΣ

**ΣΧΟΛΗ ΘΕΤΙΚΩΝ ΕΠΙΣΤΗΜΩΝ
ΤΜΗΜΑ ΠΛΗΡΟΦΟΡΙΚΗΣ ΜΕ ΕΦΑΡΜΟΓΕΣ ΣΤΗ
ΒΙΟΪΑΤΡΙΚΗ**

**Συσχέτιση miRNAs με καρκινικούς φαινοτύπους μέσω ανάλυσης
δεδομένων από CRISPR-Cas9 Knockout screens**

Βασιλική Πελεκάνου

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

**Επιβλέπουσα Καθηγήτρια:
Καθ. Άρτεμις Χατζηγεωργίου**

Λαμία, 2023

Με ατομική μου ευθύνη και γνωρίζοντας τις κυρώσεις ⁽¹⁾, που προβλέπονται από της διατάξεις της παρ. 6 του άρθρου 22 του Ν. 1599/1986, δηλώνω ότι:

1. Δεν παραθέτω κομμάτια βιβλίων ή άρθρων ή εργασιών άλλων αυτολεξεί **χωρίς να τα περικλείω σε εισαγωγικά** και χωρίς να αναφέρω το συγγραφέα, τη χρονολογία, τη σελίδα. Η αυτολεξεί παράθεση χωρίς εισαγωγικά χωρίς αναφορά στην πηγή, είναι λογοκλοπή. Πέραν της αυτολεξεί παράθεσης, λογοκλοπή θεωρείται και η παράφραση εδαφίων από έργα άλλων, συμπεριλαμβανομένων και έργων συμφοιτητών μου, καθώς και η παράθεση στοιχείων που άλλοι συνέλεξαν ή επεξεργάστηκαν, χωρίς αναφορά στην πηγή. Αναφέρω πάντοτε με πληρότητα την πηγή κάτω από τον πίνακα ή σχέδιο, όπως στα παραθέματα.
2. Δέχομαι ότι η αυτολεξεί **παράθεση χωρίς εισαγωγικά**, ακόμα κι αν συνοδεύεται από αναφορά στην πηγή σε κάποιο άλλο σημείο του κειμένου ή στο τέλος του, είναι αντιγραφή. Η αναφορά στην πηγή στο τέλος π.χ. μιας παραγράφου ή μιας σελίδας, δεν δικαιολογεί συρραφή εδαφίων έργου άλλου συγγραφέα, έστω και παραφρασμένων, και παρουσίασή τους ως δική μου εργασία.
3. Δέχομαι ότι υπάρχει επίσης περιορισμός στο μέγεθος και στη συχνότητα των παραθεμάτων που μπορώ να εντάξω στην εργασία μου εντός εισαγωγικών. Κάθε μεγάλο παράθεμα (π.χ. σε πίνακα ή πλαίσιο, κλπ), προϋποθέτει ειδικές ρυθμίσεις, και όταν δημοσιεύεται προϋποθέτει την άδεια του συγγραφέα ή του εκδότη. Το ίδιο και οι πίνακες και τα σχέδια
4. Δέχομαι όλες τις συνέπειες σε περίπτωση λογοκλοπής ή αντιγραφής.

Ημερομηνία:/...../20.....

Ο – Η Δηλ.

(Υπογραφή)

(1) «Όποιος εν γνώσει του δηλώνει ψευδή γεγονότα ή αρνείται ή αποκρύπτει τα αληθινά με έγγραφη υπεύθυνη δήλωση του άρθρου 8 παρ. 4 Ν. 1599/1986 τιμωρείται με φυλάκιση τουλάχιστον τριών μηνών. Εάν ο υπαίτιος αυτών των πράξεων σκόπευε να προσπορίσει στον εαυτόν του ή σε άλλον περιουσιακό όφελος βλάπτοντας τρίτον ή σκόπευε να βλάψει άλλον, τιμωρείται με κάθειρξη μέχρι 10 ετών.

**Συσχέτιση miRNAs με καρκινικούς φαινοτύπους μέσω ανάλυσης
δεδομένων από CRISPR-Cas9 Knockout screens**

Βασιλική Πελεκάνου

Τριμελής Επιτροπή:

Άρτεμις Χατζηγεωργίου, Καθηγήτρια (Επιβλέπουσα Καθηγήτρια)

Παντελεήμων Μπάγκος, Καθηγητής

Γεωργία Μπράλιου, Επίκουρος Καθηγήτρια

ΠΕΡΙΛΗΨΗ

Οι δοκιμασίες διαλογής ολόκληρου του γονιδιώματος με CRISPR Cas9 αποσιώπηση (genome-wide CRISPR Cas9 knockout screens) έχουν ανέλθει ως μια ισχυρή τεχνική για την ανάδειξη γονιδίων με κρίσιμους ρόλους (θετική/αρνητική επιλογή) σε μελετούμενους μηχανισμούς και φαινοτύπους. Τα microRNAs (miRNAs) είναι μικρά μη κωδικά RNAs που ρυθμίζουν μετα-μεταγραφικά την έκφραση των γονιδίων-στόχων τους. Η αιτιακή συσχέτιση της δράσης ενός miRNA με έναν μοριακό μηχανισμό αποτελεί σημαντική πρόκληση, καθώς ένα miRNA είναι ικανό να στοχεύει πολλά γονίδια και πολλά miRNAs μπορούν να ρυθμίζουν ταυτόχρονα το ίδιο γονίδιο. Στόχος της παρούσας εργασίας αποτέλεσε η ανάπτυξη μίας τεχνικής για τη βιοπληροφορική ανάλυση πειραμάτων με CRISPR Cas9 αποσιώπηση, με προσανατολισμό στα miRNAs. Με την αξιοποίηση του εργαλείου MAGeCK (Model based Analysis of Genome-wide CRISPR/Cas9 Knockout), αναλύθηκαν δεδομένα αλληλούχησης από τρεις δοκιμασίες διαλογής ολόκληρου του γονιδιώματος με CRISPR Cas9 αποσιώπηση σε καρκινικά κύτταρα: 1) ηπατοκυτταρικού καρκινώματος με αποσιώπηση του γονιδίου ATRX, 2) γαστρο-οισοφαγικού αδενοκαρκινώματος με θεραπεία Lapatinib και 3) τραχήλου της μήτρας με θεραπεία Olaparib. Τα ευρήματα που προσκομίζονται αφορούν την ανάδειξη των αλληλεπιδράσεων κρίσιμων γονιδίων και miRNAs καθώς και των ρόλων των miRNAs σε κρίσιμες βιολογικές διεργασίες.

ABSTRACT

Genome-wide CRISPR Cas9 knockout screens have emerged as a powerful technique for the identification of essential genes (positive/negative selected) in studied mechanisms and phenotypes. MicroRNAs (miRNAs) are small non-coding RNAs that regulate the expression of their target genes post-transcriptionally. The causal association between a miRNA's action and a molecular mechanism remains a significant challenge, as one miRNA can target multiple genes and multiple miRNAs can simultaneously regulate the same gene. The aim of this thesis is to develop a miRNA-oriented technique for bioinformatic analysis of CRISPR Cas9 knockout experiments. By using the MAGeCK (Model based Analysis of Genome-wide CRISPR/Cas9 Knockout) tool, sequencing data of three genome-wide CRISPR Cas9 knockout screens in cancer cells were analysed: 1) hepatocellular carcinoma cells with ATRX knockout, 2) oesophageal adenocarcinoma cells with Lapatinib treatment and 3) cervical cancer cells with Olaparib treatment. The findings involve the identification of the interactions of essential genes and miRNAs and the roles of miRNAs in essential biological processes.

Πίνακας Περιεχομένων

ΠΕΡΙΛΗΨΗ	6
ABSTRACT	7
1 Εισαγωγή	9
1.1 Η τεχνολογία CRISPR	9
1.2 Μονοπάτια επιδιόρθωσης του DNA	10
1.3 Λειτουργία και εφαρμογές των CRISPR Cas9 Knockout screens	11
1.4 Model based Analysis of Genome-wide CRISPR/Cas9 Knockout (MAGeCK)	12
1.5 Δοκιμασίες διαλογής θετικής/αρνητικής επιλογής	14
1.6 Τα microRNAs και ο ρόλος τους	14
1.7 Intragenic miRNAs και τα host γονίδιά τους	16
1.8 Εμβαθύνοντας στα miRNAs και στις αλληλεπιδράσεις με γονίδια	17
1.9 Το γονίδιο ATRX	17
1.10 Θεραπευτικός ρόλος των αναστολέων PARP στον καρκίνο	18
1.11 Καρκινική θεραπεία που στοχεύει σε υποδοχείς επιδερμικών αυξητικών παραγόντων	19
2 Υλικά και Μέθοδοι	20
2.1 Σύνολα δεδομένων CRISPR Knockout δοκιμασιών	20
2.2 Έλεγχος ποιότητας	21
2.3 Επεξεργασία δεδομένων	21
2.4 Εκτέλεση ανάλυσης MAGeCK	22
2.5 Εκτέλεση ξεχωριστών αναλύσεων των miRNA και των γονιδιακών εγγραφών και σύγκριση αποτελεσμάτων	23
2.6 Σύγκριση των βαθμολογιών των intragenic miRNAs με αυτές των host γονιδίων τους	24
2.7 Εξειδικευμένη μελέτη χαρακτηριστικών των εγγραφών για τα miRNAs και τους στόχους τους	24
3 Αποτελέσματα - Συζήτηση	28
3.1 Σύγκριση αναλύσεων miRNAs και γονιδίων από κοινού και χωριστά	28
3.2 Έλεγχος συσχέτισης βαθμολογιών των intragenic miRNAs και των γονιδίων τους	31
3.3 Κατανομή των βαθμολογιών των γονιδίων-στόχων ανά miRNA	33
3.4 Συνδυαστική ανάλυση σημαντικά επιλεγμένων miRNAs και στόχων	38
3.5 Κοινά εμπλουτισμένα μονοπάτια KEGG από επιλεγμένα miRNAs και γονίδια	44
4 Συμπεράσματα	56
Αναφορές	57

1 Εισαγωγή

1.1 Η τεχνολογία CRISPR

Το 1987, Ιάπωνες επιστήμονες ανακάλυψαν μερικές άγνωστες διαδοχικές επαναλήψεις στο γονιδίωμα του *Escherichia coli* αλλά δεν διερεύνησαν περαιτέρω τη βιολογική τους σημασία. Η έρευνα σχετικά με τις συσσωρευμένες, τακτικά παρεμβαλλόμενες, βραχείες παλινδρομικές επαναλήψεις (CRISPRs) ξεκίνησε στα μέσα της δεκαετίας του 2000, όταν ανιχνεύτηκαν σε πολλά βακτήρια και αρχαία (*archaea*)^[1]. Έπειτα, αποδείχτηκε ότι πολλά βακτήρια και τα περισσότερα αρχαία έχουν αναπτύξει καθοδηγούμενα από RNA, προσαρμοστικά, ανοσοποιητικά συστήματα, που κωδικοποιούνται από τις περιοχές CRISPR και από σχετικές συνοδευτικές πρωτεΐνες (Cas), για να τους παρέχουν επίκτητη ανοσία έναντι μόλυνσης από βακτηριοφάγους και πλασμίδια^[2]. Οι λειτουργικοί τόποι CRISPR περιλαμβάνουν μια συστοιχία (array) CRISPR πανομοιότυπων επαναλήψεων, που παρεμβάλλονται από ενδιάμεσες ακολουθίες (spacers) που στοχεύουν ξένο DNA και ένα οπερόνιο γονιδίων cas που κωδικοποιούν τα συστατικά της πρωτεΐνης Cas^[1]. Κατά τη διαδικασία ανοσοποίησης, μετά από έκθεση σε εισβάλλοντα γενετικά στοιχεία από φάγο ή πλασμίδιο, βραχεία θραύσματα από το ξένο DNA ενσωματώνονται ως νέοι spacers στους τόπους CRISPR, παρέχοντας έτσι ένα γενετικό «αρχείο» προηγούμενων μολύνσεων που επιτρέπει στον ξενιστή να αποτρέψει μελλοντική εισβολή του ίδιου εισβολέα^[2]. Ακολούθως, κατά τη διαδικασία έκφρασης, η συστοιχία CRISPR μεταγράφεται σε ένα μακρύ πρόδρομο CRISPR RNA (pre-crRNA), το οποίο επεξεργάζεται με ενδονουκλεολυτική διάσπαση έχοντας ως τελικό προϊόν το ώριμο CRISPR RNA (crRNA). Το crRNA αποτελείται από έναν μόνο spacer και από επαναλαμβανόμενες ακολουθίες CRISPR στο ένα ή και στα δύο άκρα του. Στο στάδιο της παρέμβασης, τα ώριμα crRNA συναρμολογούνται με πρωτεΐνες Cas σε σύμπλοκα με σκοπό τη στόχευση, διάσπαση και αποικοδόμηση της ειδικής συμπληρωματικής αλληλουχίας DNA^[3].

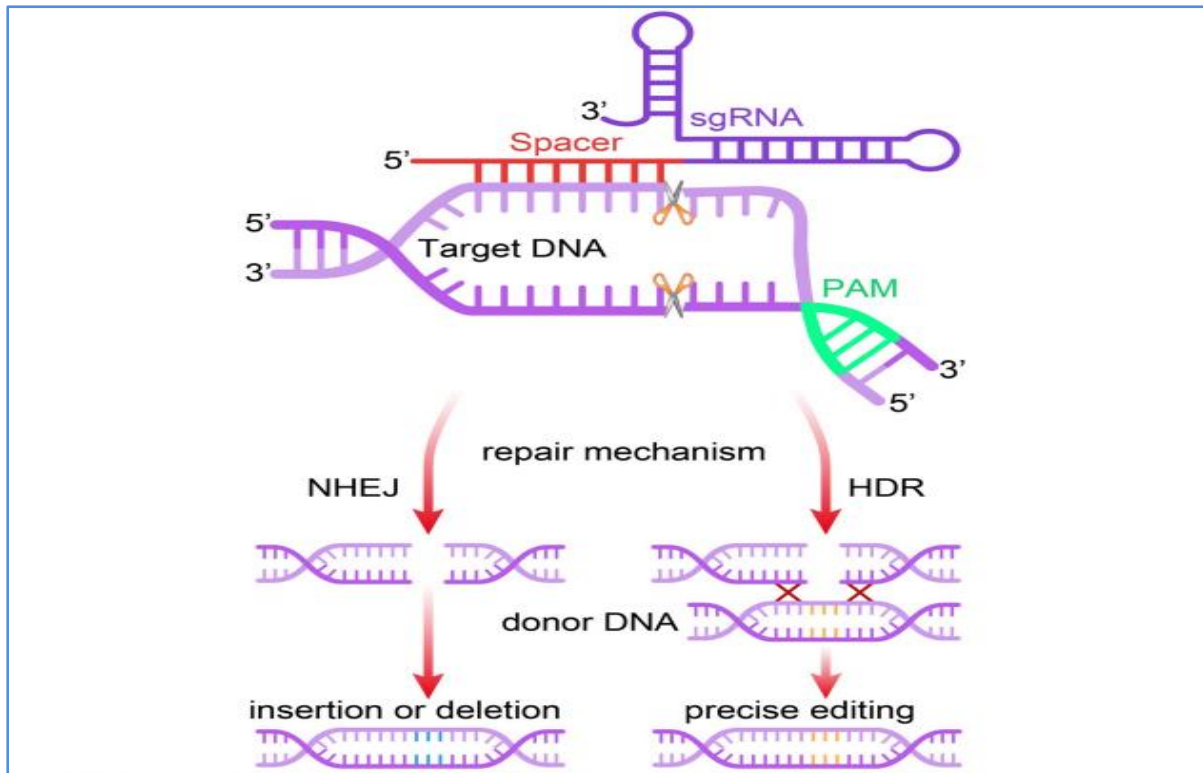
Τα συστήματα CRISPR, με βάση την οργάνωση των τόπων CRISPR και τις πρωτεΐνες Cas που περιλαμβάνουν, έχουν ταξινομηθεί σε τρεις κύριους τύπους (I, II και III), οι οποίοι υποδιαιρούνται περαιτέρω σε 11 υποτύπους^[3]. Σε αντίθεση με τα συστήματα τύπου I και τύπου III, που χρησιμοποιούν ένα μεγάλο σύμπλεγμα πολλαπλών πρωτεϊνών Cas, τα συστήματα τύπου II χρησιμοποιούν μια ενιαία ενδονουκλεάση DNA, την Cas9, για την αναγνώριση των υποστρωμάτων διπλής έλικας dsDNA και κόβουν κάθε έλικα με μία διακριτή περιοχή με δραστηριότητα νουκλεάσης (HNH ή RuvC). Κατά τη διάρκεια αυτής της διαδικασίας σίγασης, ένα επιπλέον μικρό μη κωδικό RNA, που ονομάζεται trans-activating crRNA (tracrRNA), ζευγαρώνει με την επαναλαμβανόμενη αλληλουχία στο crRNA για να σχηματίσει μια μοναδική υβριδική δομή διπλού RNA. Αυτό το δίκλωνο RNA δρα ως οδηγός κατευθύνοντας την Cas9 να διασπάσει οποιοδήποτε DNA που περιέχει συμπληρωματική αλληλουχία-στόχο (20nt) και ένα σύντομο (2-5bp), καλά συντηρημένο, παρακείμενο μοτίβο (Protospacer Adjacent Motif, PAM). Αξίζει να σημειωθεί ότι το tracrRNA απαιτείται και για την ωρίμανση του crRNA στα συστήματα CRISPR τύπου II. Ένα σύνθετο μονόκλωνο RNA

οδηγός (single guide RNA, sgRNA), που συνδυάζει το crRNA και το tracrRNA σε ένα μόνο μετάγραφο RNA, απλοποιεί το σύστημα διατηρώντας πλήρως λειτουργική τη διάσπαση της ειδικής αλληλουχίας DNA από την Cas9. Έτσι, με την αλλαγή της αλληλουχίας του guide RNA στο σημείο που βρίσκεται ο spacer, εντός του crRNA, αυτό το απλοποιημένο CRISPR-Cas9 σύστημα μπορεί να προγραμματιστεί ώστε να στοχεύει οποιαδήποτε αλληλουχία DNA στο γονιδίωμα και να δημιουργεί θραύση διπλής έλικας (Double Strand Break, DSB) με λεία άκρα (blunt end) σε συγκεκριμένο σημείο^[2].

Εν κατακλείδι, το σύστημα CRISPR-Cas9 (Clustered Regularly Interspaced Short Palindromic Repeats), αποτελεί σήμερα ένα ισχυρό εργαλείο στα χέρια του ανθρώπου για εφαρμογές γενετικής μηχανικής.

1.2 Μονοπάτια επιδιόρθωσης του DNA

Κατά τη θραύση της διπλής έλικας του DNA, από την Cas9, η στοχευόμενη περιοχή υφίσταται συνήθως τη δράση ενός από τα δύο κύρια μονοπάτια αποκατάστασης βλαβών του DNA: της, χαμηλής αξιοπιστίας, μη ομόλογης σύνδεσης άκρων (Non-homologous end joining, NHEJ), ή του μηχανισμού επιδιόρθωσης που κατευθύνεται από ομολογία (Homology directed repair, HDR) που είναι υψηλής ακρίβειας επιδιορθωτικός μηχανισμός. Κατά το NHEJ μονοπάτι, τα σπασμένα άκρα συνδέονται ξανά, χωρίς την ανάγκη να υπάρχει ομόλογο πρότυπο, με αποτέλεσμα συχνά να προκύπτουν μεταλλάξεις προσθαφαιρέσεων (insertions/deletions, indels). Έτσι, το NHEJ, μπορεί να αξιοποιηθεί στη γονιδιακή επεξεργασία ως μεσολαβητής για εξάλειψη γονιδίων (gene knockouts), καθώς τα indels εντός ενός κωδικού εξονίου μπορεί να οδηγήσουν σε μεταλλάξεις μετατόπισης αναγνωστικού πλαισίου (frameshift) και σε πρόωρα κωδικόνια τερματισμού (premature stop codons), με αποτέλεσμα τη γονιδιακή σίγαση. Το HDR είναι ένα εναλλακτικό κύριο μονοπάτι επιδιόρθωσης του DNA^[4]. Αξιοποιεί ομόλογες αλληλουχίες DNA προερχόμενες από αδελφές χρωματίδες, ομόλογα χρωμοσώματα ή εξωγενή μόρια DNA για την παραγωγή ακριβών εισαγωγών, διαγραφών ή υποκαταστάσεων βάσεων σε μια θέση DSB ή μεταξύ δύο DSBs^[5]. Παρόλο που το HDR εμφανίζεται συνήθως με μικρότερη και αρκετά μεταβλητή συχνότητα μπορεί να αξιοποιηθεί για να δημιουργήσει ακριβείς, καθορισμένες τροποποιήσεις σε έναν τόπο-στόχο. Σε αντίθεση με το NHEJ, το HDR είναι ενεργό μόνο σε διαιρούμενα κύτταρα και η αποτελεσματικότητά του ποικίλλει ανάλογα με την κατάσταση και τον τύπο του κύτταρου^[4]. Καθένα από τα δύο αυτά μονοπάτια μπορεί να αξιοποιηθεί σε εφαρμογές γενετικής μηχανικής σε συνδυασμό με την τεχνική CRISPR Cas9, ανάλογα με τις ανάγκες του πειράματος (Εικόνα 1).



Εικόνα 1: Ο μηχανισμός γονιδιακής επεξεργασίας με CRISPR Cas9. Το sgRNA κατευθύνει τη νουκλεάση Cas9 σε μια συμπληρωματική αλληλουχία στο γονιδίωμα όπου η Cas9 θα προκαλέσει DSB. Ο γονιδιωματικός τόπος-στόχος πρέπει να ακολουθείται από ένα μοτίβο PAM για να λειτουργήσει η Cas9. Τα DSBs επισκευάζονται με ένα από τα δυο κύρια μονοπάτια επιδιόρθωσης του DNA: Το NHEJ που συχνά οδηγεί σε gene knockouts μέσω indels και το HDR το οποίο μπορεί να αξιοποιηθεί για την εισαγωγή ακριβών γενετικών τροποποιήσεων ή εξωγενών αλληλουχιών^[6].

1.3 Λειτουργία και εφαρμογές των CRISPR Cas9 Knockout screens

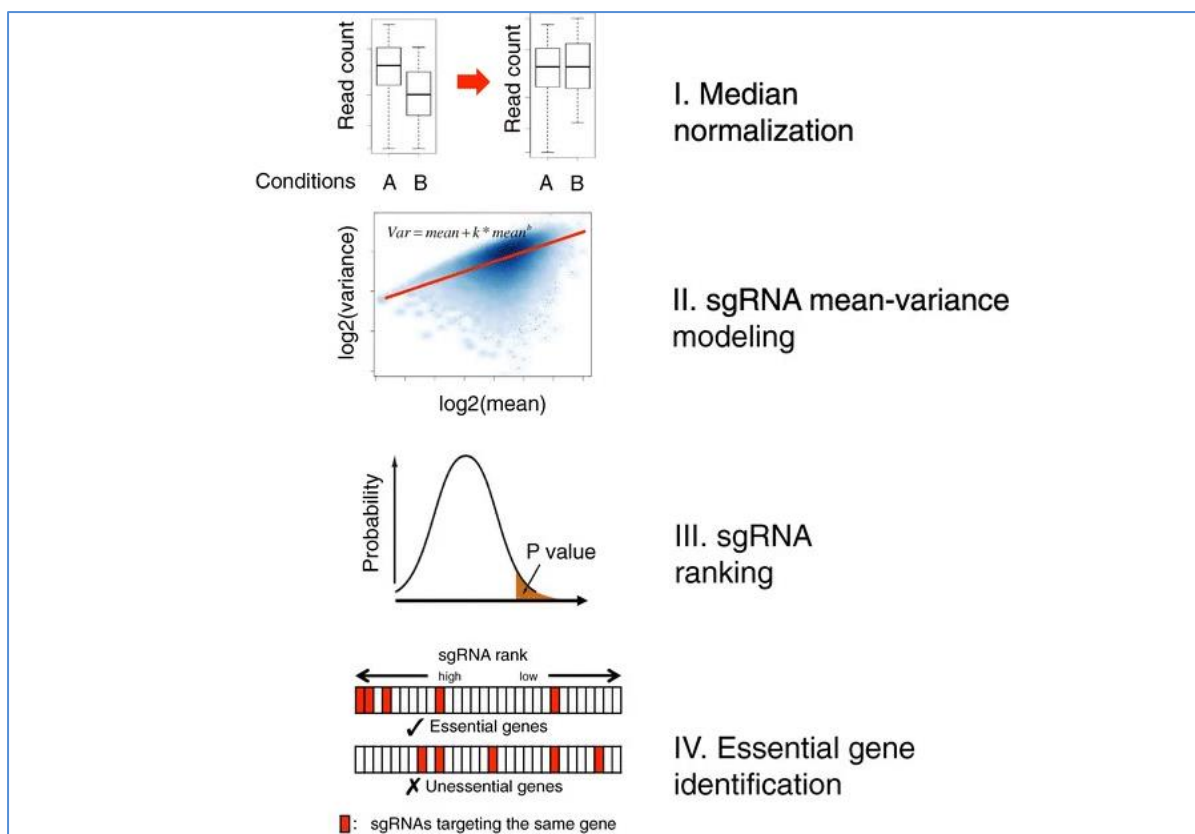
Οι γενετικές δοκιμασίες διαλογής (genetic screens) αποτελούν μια μέθοδο με πρωταρχικό σκοπό την προσέγγιση της σύνδεσης-συσχέτισης μεταλλαγμένων φαινοτύπων με τη γενετική τους βάση. Συγκεκριμένα, οι οθόνες CRISPR αποτελούν μια μεγάλης κλίμακας πειραματική προσέγγιση ελέγχου, ενός πληθυσμού κυττάρων για την ανακάλυψη γονιδίων, που εμπλέκονται σε έναν συγκεκριμένο φαινότυπο. Η αποτελεσματικότητα της ενδονουκλεάσης Cas9 και η ευκολία με την οποία συγκεκριμένα γονίδια μπορούν να στοχευθούν και να σιγηθούν, απλώς αλλάζοντας την αλληλουχία του οδηγού-RNA (sgRNA), έχουν οδηγήσει στη δημιουργία CRISPR-knockout (KO) βιβλιοθηκών από sgRNAs, οι οποίες είναι εμπορικά διαθέσιμες και έχουν εύρος που καλύπτει ολόκληρο το γονιδίωμα (genome-wide). Αυτές οι βιβλιοθήκες είναι συνήθως βελτιστοποιημένες για την επίτευξη ίσης εκπροσώπησης και απόδοσης σε όλα τα εκφραζόμενα sgRNAs και φέρουν ένα αντιβιοτικό ή δείκτη επιλογής φθορισμού, που μπορεί να χρησιμοποιηθεί για την ανάκτηση των κυττάρων που έχουν μετασηματιστεί επιτυχώς. Μετά από την επιλογή αντιβιοτικού (antibiotic selection) και την ανάπτυξη και/ή εφαρμογή της sgRNA βιβλιοθήκης, τα κύτταρα μπορούν να

υποβληθούν σε πίεση επιλογής (selection pressure), η οποία έχει ως αποτέλεσμα να μεταβληθεί η σύσταση του πληθυσμού των μετασχηματισμένων κυττάρων και να εμπλουτίζεται σε κύτταρα που φέρουν συγκεκριμένα ωφέλιμα sgRNAs. Στη συνέχεια, μέσα από (α) απομόνωση του DNA, (β) ενίσχυση με PCR και (γ) Αλληλούχηση Επόμενης Γενιάς (Next Generation Sequencing, NGS) και (δ) την βιοπληροφορική ανάλυση, είναι δυνατή η ποσοτικοποίηση κάθε sgRNA, καθώς και η συσχέτισή του με γονίδια-στόχους. Υπάρχουν αρκετά βιοπληροφορικά εργαλεία που μπορούν να χρησιμοποιηθούν για την αξιολόγηση της στατιστικής σημασίας των πιο εμπλουτισμένων ή εξαντλημένων (depleted) sgRNAs και γονιδίων, με ένα από αυτά που χρησιμοποιείται ευρέως να είναι το πακέτο MAGeCK^[7].

1.4 Model based Analysis of Genome-wide CRISPR/Cas9 Knockout (MAGeCK)

Το MAGeCK αποτελεί ένα υπολογιστικό εργαλείο για τον εντοπισμό απαραίτητων (essential) sgRNAs, γονιδίων και βιολογικών μονοπατιών, από τις CRISPR Cas9 Knockout δοκιμασίες. Υπερέχει από τις υπόλοιπες σχετικές υπολογιστικές μεθόδους επειδή (α) εφαρμόζει έλεγχο του σφάλματος τύπου I (εσφαλμένη απόρριψη της μηδενικής υπόθεσης) με τη μέθοδο False Discovery Rate (FDR), (β) για την υψηλή ευαισθησία του, (γ) για τη διατήρηση ισχυρών αποτελεσμάτων σε διαφορετικά βάθη αλληλούχησης και πλήθη των sgRNAs ανά γονίδιο, καθώς και (δ) για το γεγονός ότι επιτρέπει την ανάδειξη σημαντικών γονιδίων που υφίστανται τόσο θετική όσο και αρνητική επιλογή από την ίδια πειραματική διάταξη^[8].

Δύο διακριτές μεθοδολογίες προσφέρονται από το MAGeCK για τον εντοπισμό σημαντικά επιλεγμένων στοιχείων, ο αλγόριθμος RRA (Robust Rank Aggregation) και ο MLE (Maximum Likelihood Estimation). Ο RRA επιτρέπει τη σύγκριση μεταξύ δύο πειραματικών συνθηκών. Κατατάσσει τα sgRNAs με βάση τα p-values που υπολογίζει χρησιμοποιώντας Negative Binomial (NB) μοντέλο για να ελέγξει αν η αφθονία των sgRNAs διαφέρει στατιστικώς σημαντικά μεταξύ των συνθηκών. Έπειτα, για να εντοπίσει τα θετικά και τα αρνητικά επιλεγμένα γονίδια, εφαρμόζει έναν τροποποιημένο αλγόριθμο RRA, ο οποίος ονομάζεται α-RRA και στηρίζεται στην ανάδειξη ασυμμετρίας (skewness) στην κατανομή των βαθμίδων των sgRNAs που στοχεύουν κάθε γονίδιο ξεχωριστά (**Εικόνα 2**). Ο MLE, από την άλλη, είναι ένας Εκτιμητής Μέγιστης Πιθανοφάνειας που μπορεί να χρησιμοποιηθεί και για την ανάλυση δεδομένων από δοκιμασίες που περιλαμβάνουν περισσότερες από δύο συνθήκες. Επίσης, αξιοποιεί πληροφορία για την αποδοτικότητα (efficiency) κάθε sgRNA, ενώ υπολογίζει ένα β score (ανάλογο του $\log_2(\text{fold-change})$) που χρησιμοποιείται σε πειράματα RNA-Seq) για κάθε στοχευόμενο γονίδιο για να μετρήσει τον βαθμό επιλογής κάθε γονιδίου^[9]. Ένα θετικό β score συμβολίζει θετική επιλογή, ενώ ένα αρνητικό αρνητική επιλογή.



Εικόνα 2: Επισκόπηση του αλγορίθμου MAGeCK. Οι ακατέργαστες μετρήσεις των εγγραφών (*Read Counts*), που αντιστοιχούν σε *sgRNAs*, από διαφορετικά πειράματα κανονικοποιούνται εφαρμόζοντας την μέθοδο *median ratio* σε όλα τα πειράματα. Έπειτα, εφαρμόζεται η συνάρτηση μέσου-διακύμανσης (*mean-variance*), η οποία καταγράφει τη σχέση της μέσης τιμής και της διακύμανσης στα αντίγραφα. Η στατιστική σημασία κάθε *sgRNA* υπολογίζεται χρησιμοποιώντας αρνητική διωνυμική κατανομή (*Negative Binomial distribution*) με παραμέτρους από το μοντέλο *mean-variance*. Ως απαραίτητα (*essential*) γονίδια (θετικά και αρνητικά επιλεγμένα) προσδιορίζονται αυτά των οποίων τα *sgRNAs* κατανέμονται σταθερά υψηλότερα (σε σημαντικότητα), ενώ ως μη απαραίτητα αυτά των οποίων τα *sgRNAs* έχουν ανομοιόμορφη κατανομή (μέθοδος *Robust Rank Aggregation, RRA*)^[8].

Στις *downstream* αναλύσεις απαιτείται και η βιολογική λειτουργική ανάλυση για την αναγνώριση των βιολογικών λειτουργιών των εντοπισμένων γονιδίων. Για αυτόν τον σκοπό, το *MAGeCKFlute* αποτελεί ένα χρήσιμο πλήρες *pipeline*, που περιέχει μια σειρά από συναρτήσεις για την ανάλυση των *CRISPR* δοκιμασιών. Μπορεί να εφαρμοστεί στα αποτελέσματα των αλγορίθμων *RRA* και *MLE*, παρέχοντας στρατηγικές (α) για την αφαίρεση πιθανών *biases* στο πλήθος των αναγνώσεων των *sgRNAs* και στα β scores των γονιδίων, (β) για την εύρεση απαραίτητων και μη απαραίτητων γονιδίων και (γ) για την εκτέλεση ανάλυσης εμπλουτισμού μονοπατιών (*Pathway Enrichment Analysis*)^[9].

1.5 Δοκιμασίες διαλογής θετικής/αρνητικής επιλογής

Οι δοκιμασίες διαλογής μπορούν να χωριστούν σε δύο τύπους: θετικής επιλογής και αρνητικής επιλογής. Αμφότεροι οι τύποι διαλογής χρησιμοποιούν μία μέθοδο επιλογής σχετικής με τον υπό μελέτη φαινότυπο.

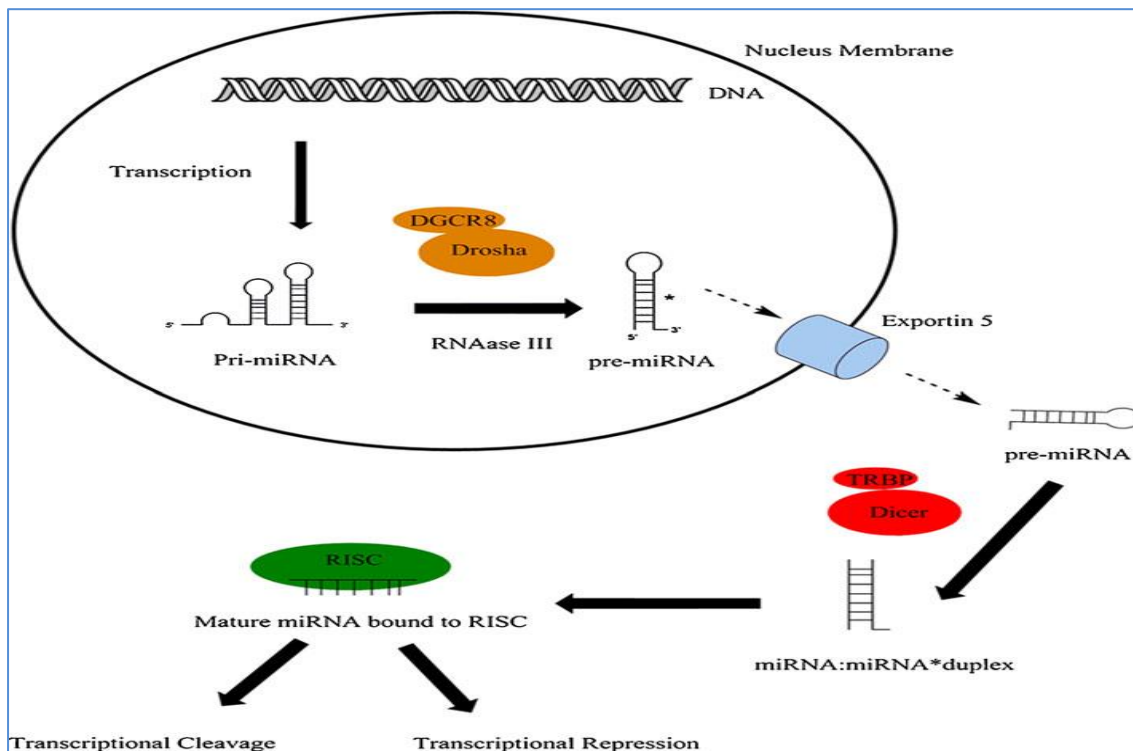
- Στις δοκιμασίες διαλογής θετικής επιλογής και κατόπιν της εφαρμογής της επιλογής, τα περισσότερα κύτταρα πεθαίνουν και εκτελείται αλληλούχηση των κυττάρων που καταφέρνουν να επιβιώσουν. Ο στόχος αυτών των δοκιμασιών είναι ο εντοπισμός των μορίων των οποίων η παρουσία/απουσία βοήθησε το κύτταρο να επιβιώσει έπειτα από την επιλογή.
- Στις δοκιμασίες αρνητικής επιλογής, συγκρίνεται ένα δείγμα ελέγχου (δεδομένα αλληλούχησης πριν την εκτέλεση της επιλογής) με ένα τελικό δείγμα (δεδομένα αλληλούχησης μετά από την εκτέλεση της επιλογής) από την κυτταρική καλλιέργεια. Σε αυτή την περίπτωση, τα περισσότερα κύτταρα επιβιώνουν μετά από την εκτέλεση της επιλογής και επιδιώκεται ο εντοπισμός των μορίων που είναι κρίσιμα (με αρνητικό ή θετικό ρόλο) για την ανάπτυξη και τον πολλαπλασιασμό των κυττάρων.

Με την αξιοποίηση του εργαλείου MAGeCK δίνεται η δυνατότητα εντοπισμού τόσο θετικά όσο και αρνητικά επιλεγμένων μορίων στις δοκιμασίες διαλογής, ανεξάρτητα από τον τύπο επιλογής, που έχει εφαρμοστεί. Οι βαθμολογίες που αποδίδονται για την θετική επιλογή υποδηλώνουν την αύξηση της αφθονίας των sgRNAs (μεταξύ των συγκρινόμενων συνθήκων), ενώ οι βαθμολογίες αρνητικής επιλογής την μείωσή της.

1.6 Τα *microRNAs* και ο ρόλος τους

Τα ώριμα *microRNAs* (miRNAs) είναι μονόκλωνα, μη κωδικά μόρια RNA, μήκους 20-23nt, που έχουν σημαντικούς ρυθμιστικούς ρόλους σε όλα τα ζωικά και φυτικά κύτταρα^[12]. Τα miRNAs μεταγράφονται στον πυρήνα του κυττάρου από αλληλουχίες DNA σε πρωτογενή miRNAs (primary miRNAs, pri-miRNAs) και υποβάλλονται σε επεξεργασία από πρωτεϊνικά σύμπλοκα για να γίνουν πρόδρομα miRNAs (precursor miRNAs, pre-miRNAs). Μόλις δημιουργηθούν, τα πρόδρομα miRNAs, εξάγονται στο κυτταρόπλασμα και υποβάλλονται σε επεξεργασία από την ενδονουκλεάση Dicer^[13]. Το προϊόν της επεξεργασίας αυτής είναι ένα δίκλωνο RNA από το οποίο ο ένας κλώνος δεσμεύεται σε ένα πολυπρωτεϊνικό σύμπλοκο σίγασης που προκαλείται από RNA (σύμπλοκο RISC), ως ώριμο miRNA, και ο άλλος κλώνος, τυπικά, αποικοδομείται. Ωστόσο, για ορισμένα miRNAs και οι δύο κλώνοι μπορεί να δεσμευθούν σε RISC, με παρόμοιες συχνότητες, ή με διαφορετική προτίμηση ανά κυτταρικό τύπο και συνθήκη. Σε αυτή την περίπτωση, ο κλώνος από το 5' άκρο της φουρκέτας ονομάζεται "5p" και ο κλώνος από το 3' άκρο ονομάζεται "3p"^[14]. Μόλις ενσωματωθεί στο RISC, το miRNA καθοδηγεί το σύμπλοκο miRISC στους mRNA-στόχους του μέσω αλληλεπιδράσεων ζευγών βάσεων, με τέλεια ή σχεδόν τέλεια συμπληρωματικότητα ^[15]. Συνήθως, η θέση-στόχος στο mRNA είναι η 3' αμετάφραστη περιοχή (3'UTR), και η

συμπληρωματική αλληλουχία στο miRNA, που ονομάζεται αλληλουχία seed, βρίσκεται στα νουκλεοτίδια 2-7 από το 5' άκρο του ^[16]. Μετά την αναγνώριση του mRNA-στόχου, είτε παρεμποδίζεται η μετάφρασή του (μέσω παρεμπόδισης της επιμήκυνσης μετάφρασης, μέσω αποικοδόμησης πρωτεϊνών ή μέσω πρόωρης αποδέσμευσης του ριβοσώματος (ribosome drop-off)), είτε οδηγείται προς αποδόμηση. Αρχικές μελέτες έδειξαν ότι εάν το ζεύγος miRNA και mRNA-στόχου περιέχει ατελή συμπληρωματική ποσότητα, η έκφραση πρωτεΐνης αναστέλλεται χωρίς να καταστρέφεται το αντίστοιχο mRNA. Ωστόσο, εάν το ζεύγος έχει σχεδόν τέλεια συμπληρωματικότητα, τότε ο mRNA-στόχος, μετά από αποαδενυλίωση και αποσύνθεση, αποδομείται^[16, 17] (Εικόνα 3).



Εικόνα 3: Η βιογένεση και επεξεργασία του miRNA. Στον πυρήνα του κυττάρου γίνεται η μεταγραφή ενός μεγάλου πρωτογενούς miRNA (*pri-miRNA*) από την RNA πολυμεράση II. Αυτό το μεγάλο *pri-miRNA* διασπάται στη συνέχεια από το ένζυμο *RNAse III Drosha* και συνδέεται με την υπομονάδα του συμπλέγματος μικροεπεξεργαστή *DGCR8* για την παραγωγή του πρόδρομου miRNA (*pre-miRNA*). Το *pre-miRNA* (70-90nt) περιέχει μια δομή βρόχου στελέχους και μεταφέρεται στο κυτταρόπλασμα του κυττάρου από την πρωτεΐνη *Exportin-5*. Μόλις εξέλθει στο κυτταρόπλασμα, αυτή η δομή φουρκέτας κόβεται από το ένζυμο *Dicer* (πρωτεΐνη της οικογένειας *RNAse III*) παράγοντας το δίκλωνο miRNA (*miRNA:miRNA duplex*). Η διαδικασία ολοκληρώνεται με την ενσωμάτωση του ώριμου κλώνου του miRNA στο σύμπλοκο *RISC*. Σε αυτόν το σχηματισμό, το *miRISC* στοχεύει συμπληρωματικές αλληλουχίες mRNA με σκοπό τη διάσπαση ή τη μεταγραφική καταστολή.^[18]

Τα miRNAs ανακαλύφθηκαν στις αρχές της δεκαετίας του 1990 από τον Victor Ambros και τους συνεργάτες του. Βρέθηκε ότι το *lin-4*, ένα γνωστό γονίδιο που εμπλέκεται στην ανάπτυξη του νηματώδους *Caenorhabditis elegans* (*C. elegans*), δεν κωδικοποιεί κάποια

πρωτεΐνη αλλά, αντίθετα, παράγει ένα μικρό RNA μήκους 22 νουκλεοτιδίων, το οποίο αποδείχθηκε ότι αλληλεπιδρά με την 3' UTR του mRNA του *lin-14* και καταστέλλει την έκφρασή του^[19]. Αργότερα, το 2000, δύο ερευνητικές ομάδες ανακάλυψαν ότι ένα μικρό RNA, το *let-7*, το οποίο βρέθηκε ότι ήταν απαραίτητο για την ανάπτυξη του *C. elegans* από το μεταγενέστερο προνυμφικό στάδιο σε ενήλικα. Στη συνέχεια, ομόλογα αυτού του γονιδίου ανακαλύφθηκαν σε πολλούς άλλους οργανισμούς, όπως και στον άνθρωπο^[20]. Έτσι, οι έρευνες πάνω στα miRNAs σταδιακά αυξήθηκαν αποδεικνύοντας εντονότερα το ρόλο τους στην αρνητική ρύθμιση της έκφρασης των mRNAs αλλά και στον έλεγχο του χρόνου ανάπτυξης των κύτταρων και κατ' επέκταση των οργανισμών. Σήμερα, είναι γνωστό ότι τα miRNAs έχουν πολύ σημαντικές ρυθμιστικές λειτουργίες σε βασικές βιολογικές διεργασίες, όπως η ανάπτυξη, η κυτταρική διαφοροποίηση, ο κυτταρικός πολλαπλασιασμός και η κυτταρική απόπτωση, οι οποίες επηρεάζουν υψηλότερου επιπέδου βιολογικές λειτουργίες όπως η διατήρηση του κυτταρικού δυναμικού διαφοροποίησης και αυτοανανέωση, η ανοσία και η καρκινογένεση. Τα miRNAs έχουν μελετηθεί εντατικά στον τομέα της ογκολογικής έρευνας και αναδυόμενα στοιχεία δείχνουν ότι αλλαγές στις ρυθμιστικές λειτουργίες των miRNAs εμπλέκονται στην παθογένεση καρκίνων, κυρίως στη ρύθμιση της μετάφρασης ογκογονιδίων και ογκοκατασταλτικών γονιδίων^[19]. Αξίζει να αναφερθεί ότι το κάθε miRNA μπορεί να ρυθμίζει πολλαπλά γονίδια, ενώ υπολογίζεται πως περισσότερο από το ένα τρίτο όλων των ανθρώπινων γονιδίων στοχεύεται από miRNAs^[21].

1.7 Intragenic miRNAs και τα host γονιδιά τους

Τα miRNAs με κριτήριο τη θέση τους στο γονιδίωμα μπορούν να ταξινομηθούν σε "intragenic" miRNAs και σε "intergenic" miRNAs. Αυτά που εντοπίζονται εντός ιντρονίων ή εξονίων γονιδίων ονομάζονται intragenic και τα γονίδια στα οποία είναι ενσωματωμένα ονομάζονται host γονίδια. Τα intergenic miRNAs εντοπίζονται ανάμεσα σε γονίδια και πιστεύεται ότι μεταγράφονται από δικούς τους υποκινητές. Ωστόσο, μελέτες έδειξαν ότι υπάρχει ένα μικρό ποσοστό intergenic miRNAs τα οποία δυνητικά μεταγράφονται μαζί με τα γειτονικά τους γονίδια. Αυτό συμβαίνει είτε όταν η μεταγραφή ενός γειτονικού γονιδίου συνεχίζεται πέρα από το σημείο λήξης μεταγράφοντας και το miRNA που βρίσκεται στην αμέσως εγγύτερα καθοδική περιοχή (readthrough transcription), είτε όταν δύο πολυμεράσες μεταγράφουν αντίστοιχα και με τις δυο κατευθύνσεις (σημαίνουσα και αντισημαίνουσα) ξεκινώντας από κοινό υποκινητή (divergent transcription). Αυτές οι δύο κατηγορίες intergenic miRNAs κατατάσσονται στην intragenic κατηγορία επειδή μοιράζονται τον ίδιο υποκινητή με τα γειτονικά τους γονίδια^[22].

Τα intragenic miRNAs μπορούν να κατηγοριοποιηθούν περαιτέρω σε ιντρονικά miRNAs (που βρίσκονται σε ιντρονικές περιοχές), σε εξονικά miRNAs (που βρίσκονται σε εξονικές περιοχές), σε "junction" miRNAs (miRNA που βρίσκονται σε συνδέσεις ιντρονίου-εξονίου) και σε αντι-νοηματικά miRNAs (miRNAs που εκφράζονται από τον αντίθετο κλώνο από αυτόν

που εκφράζεται το γονίδιο)^[22]. Ένας σημαντικός αριθμός αυτών των miRNAs συμμεταγράφονται και συρρυθμίζονται με τα host γονιδιά τους^[23].

1.8 Εμβαθύνοντας στα miRNAs και στις αλληλεπιδράσεις με γονίδια

Τα διαρκώς αυξανόμενα δεδομένα που προκύπτουν από έρευνες για τα miRNAs ταξινομούνται σε διάφορους τομείς, όπως η ανακάλυψη νέων miRNAs, οι αλληλουχίες των ώριμων και πρόδρομων miRNAs, ο ρόλος τους σε ασθένειες, οι αλληλεπιδράσεις τους σε μονοπάτια, οι προβλέψεις των στόχων τους κ.λπ. Για να προσεγγιστούν αυτές οι πληροφορίες, έχουν δημιουργηθεί βάσεις δεδομένων και λογισμικά-εργαλεία εξειδικευμένα στα miRNAs^[24]. Κάποια από αυτά είναι: α) η βάση δεδομένων miRBase, β) η βάση δεδομένων DIANA-TarBase, γ) ο αλγόριθμος DIANA-microT-CDS και δ) το λογισμικό DIANA-miRPath. Η miRBase είναι ο κύριος δημόσιος χώρος αποθήκευσης και διαδικτυακός πόρος για miRNA αλληλουχίες. Για κάθε miRNA που είναι καταχωρημένο, η miRBase περιέχει την αλληλουχία του, το πρόδρομο miRNA από το οποίο γίνεται η βιογένεσή του, τις συντεταγμένες του στο γονιδίωμα, βιβλιογραφικές αναφορές, δεδομένα έκφρασης από πειράματα αλληλούχησης υψηλής-απόδοσης και αναφορές σχετικά με την ανακάλυψή του^[25]. Η DIANA-TarBase v8 είναι μια βάση δεδομένων εξειδικευμένη στην εύρεση στόχων των miRNAs σε κώδικά και σε μη κωδικά RNA, οι οποίοι υποστηρίζονται πειραματικά^[26]. Παρέχει, επίσης, πρόσβαση σε δημοσιεύσεις σχετικά με την κάθε αλληλεπίδραση και σε περισσότερες πληροφορίες για τα miRNAs και τα γονίδια (τον κυτταρικό τύπο στον οποίο εκφράζονται, συσχετιζόμενα μονοπάτια κ.λπ.). Ο αλγόριθμος DIANA-microT-CDS είναι ειδικά σχεδιασμένος για την αναγνώριση των στόχων των miRNAs τόσο στις 3'UTRs όσο και στις κωδικές περιοχές^[27]. Το miRPath είναι ένα διαδικτυακό λογισμικό εξειδικευμένο στην εκτίμηση των ρυθμιστικών ρόλων των miRNAs και στον εντοπισμό των μονοπατιών που ελέγχουν. Καθιστά δυνατή την επισήμανση της λειτουργίας ενός ή περισσότερων miRNAs, χρησιμοποιώντας τυπικές κατανομές (υπεργεωμετρικές κατανομές), αμερόληπτες (unbiased) εμπειρικές κατανομές ή/και στατιστικές μεθοδολογίες μετα-ανάλυσης. Επιπλέον, έχει επεκταθεί σημαντικά, ώστε να υποστηρίζει αναλύσεις για μονοπάτια KEGG^[28] (Kyoto Encyclopedia of Genes and Genomes), GO^[29] (GeneOntology), Reactome^[30], MSigDB^[31] (Molecular Signatures Database)^[32].

1.9 Το γονίδιο ATRX

Το γονίδιο ATRX ανακαλύφθηκε σε ασθενείς με X-linked mental retardation syndrome (σύνδρομο ATRX) και βρίσκεται στο X χρωμόσωμα^[33]. Βρέθηκε ότι κωδικοποιεί μια πρωτεΐνη αναδιαμόρφωσης χρωματίνης (μέλος της οικογένειας SWI2/SNF2), η οποία δρα ως μεταγραφικός ρυθμιστής, με κύρια λειτουργία τη δέσμευση ιστόνης H3.3 (εναλλακτική μορφή της ιστόνης H3) σε ακολουθίες τελομερών και άλλες γονιδιωματικές περιοχές με επαναλαμβανόμενες αλληλουχίες (heterochromatic repeats)^[34]. Για τη δέσμευση και

εναπόθεση της ιστόνης H3.3, η ATRX πρωτεΐνη αλληλεπιδρά με το συμπαράγοντα μεταγραφής DAXX (death domain associated protein). Αυτή η βιολογική διεργασία καταλαμβάνει σημαντικό ρόλο στη συναρμολόγηση και σταθεροποίηση της χρωματίνης [35].

Μεταλλάξεις στο ATRX έχουν συσχετιστεί με τον εναλλακτικό μηχανισμό επιμήκυνσης των τελομερών (Alternative Lengthening of Telomeres, ALT). Η απώλεια ATRX έχει επίσης συνδεθεί στενά με βλάβη στο DNA και με replicative stress (κυτταρικό στρες σχετιζόμενο με την αντιγραφή του DNA). Οι παράγοντες που προκαλούν απουσία έκφρασης του ATRX είναι συνήθως μεταλλάξεις, διαγραφές ή συγχωνεύσεις γονιδίων^[36]. Πρόσφατες γονιδιωματικές μελέτες έχουν δείξει ότι το ATRX συναντάται σε μεταλλαγμένες μορφές σε πολλούς ανθρώπινους καρκίνους (νευροενδοκρινής όγκος παγκρέατος, γλοίωμα, καρκίνος του ήπατος και νευροβλάστωμα). Οι σωματικές μεταλλάξεις του ATRX, που σχετίζονται με την εμφάνιση καρκίνου, έχουν συνήθως ως αποτέλεσμα την ελλιπή έκφρασή του, υποδεικνύοντας έτσι ότι το ATRX λειτουργεί ως ογκοκατασταλτικό γονίδιο σε αυτούς τους καρκίνους^[37]. Τα ευρήματα αυτά εντείνουν το ενδιαφέρον για περαιτέρω μελέτη του μεταλλαγμένου ATRX καθώς και των καρκινικών κυττάρων που φέρουν μετάλλαξη ATRX. Παρόλο που με την ανάπτυξη των τεχνικών αλληλούχησης αναδεικνύονται γρήγορα και αξιόπιστα μεταλλάξεις που οδηγούν στην εμφάνιση καρκίνου, παραμένει σημαντική η πρόκληση για αποτελεσματική μετάφραση και χρήση αυτών των δεδομένων σε βιώσιμες θεραπείες.

1.10 Θεραπευτικός ρόλος των αναστολέων PARP στον καρκίνο

Τα ένζυμα Poly (ADP-ribose) polymerase (PARP) συνιστούν μια οικογένεια πρωτεϊνών (PARP family), που τα μέλη της χαρακτηρίζονται από σημαντική δομική ομοιότητα μεταξύ τους, και παρόμοιες λειτουργίες^[38]. Τα ένζυμα αυτά, διαδραματίζουν σημαντικό ρόλο σε κυτταρικές διεργασίες, όπως η γονιδιακή ρύθμιση, η αναδιαμόρφωση της χρωματίνης, η επιδιόρθωση του DNA και η απόπτωση^[39]. Το PARP-1 είναι το πρώτο χαρακτηρισμένο και εκτενώς μελετημένο ένζυμο από την οικογένεια PARP, που αναγνωρίστηκε ότι παίζει ουσιαστικό ρόλο στην επιδιόρθωση του DNA και στην κυτταρική απόπτωση^[38, 39]. Το PARP-1 χρησιμοποιεί μόρια NAD⁺ (δινουκλεοτίδιο νικοτιναμίδης αδενίνης) για την κατάλυση της σύνθεσης poly ADP-ριβόζης (PAR), με την απελευθέρωση νικοτιναμίδης και μεταφέρει τμήματα από αυτή σε πρωτεΐνες-δέκτες. Με τη δέσμευση του PARP-1 σε περιοχές DNA με βλάβη επηρεάζονται οι πρωτεΐνες που εμπλέκονται στον μεταβολισμό του DNA. Η προσθήκη τμημάτων poly ADP-ριβόζης οδηγεί στην αναδιαμόρφωση της χρωματίνης και στην κινητοποίηση των πρωτεϊνών επισκευής του DNA για την εξάλειψη βλαβών^[40]. Το γεγονός ότι το PARP1 είναι ένα βασικό ένζυμο επιδιόρθωσης του DNA και σε καρκινικά κύτταρα, έστρεψε την προσοχή στην αναζήτηση αναστολέων PARP για χρήση σε καρκινικές θεραπείες.

Η αναστολή της επιδιόρθωσης του DNA στα καρκινικά κύτταρα αντιπροσωπεύει μια ελκυστική στρατηγική για την ενίσχυση των κυτταροτοξικών επιδράσεων της

χημειοθεραπείας και της ακτινοβολίας και ως εκ τούτου αποτελεί αντικείμενο επιστημονικής έρευνας για αρκετές δεκαετίες. Οι αναστολείς PARP (PARPi) είναι οι πιο αναπτυγμένοι και φαίνονται πολλά υποσχόμενοι σε μια ποικιλία τύπων καρκίνου, συμπεριλαμβανομένων των καρκίνων του μαστού και των ωοθηκών^[41]. Πιο συγκεκριμένα, αποδείχθηκαν ιδιαίτερα τοξικοί για κύτταρα με έλλειψη, ομόλογου ανασυνδυασμού (Homologous Recombination, HR), σε BRCA1 (πρωτεΐνη ευαισθησίας για καρκίνο του μαστού τύπου 1) και BRCA2 (πρωτεΐνη ευαισθησίας για καρκίνο του μαστού τύπου 2), ακόμα και απουσία εξωγενούς βλάβης DNA, φαινόμενο με μεγάλη θεραπευτική προοπτική, λόγω του υψηλού επιπολασμού της ανεπάρκειας BRCA στα καρκινικά κύτταρα^[42]. Το Olaparib ήταν το πρώτο PARPi μόριο που εγκρίθηκε για τη θεραπεία του καρκίνου μαστού και ωοθηκών με έλλειψη BRCA και αναπτύχθηκε από την εταιρεία AstraZeneca^[38, 42, 43].

1.11 Καρκινική θεραπεία που στοχεύει σε υποδοχείς επιδερμικών αυξητικών παραγόντων

Οι υποδοχείς επιδερμικών αυξητικών παραγόντων στον άνθρωπο (HER, EGFR, ErbB) είναι μια οικογένεια πρωτεϊνών που αποτελείται από τέσσερις υποδοχείς κινάσες τυροσίνης (Receptor Tyrosine Kinases, RTKs): HER1, HER2, HER3 και HER4. Οι υποδοχείς κινάσες τυροσίνης είναι διαμεμβρανικές πρωτεΐνες που διαθέτουν μια περιοχή δέσμησης, στην εξωκυτταρική τους πλευρά, με συγκεκριμένες θέσεις πρόσδεσης για διάφορες πρωτεΐνες προσαρμογής (adaptor proteins) και συνδέτες (ligands), μια διαμεμβρανική περιοχή και μια ενδοκυτταρική κυτταροπλασματική περιοχή που περιέχει την καταλυτική θέση κινάσης τυροσίνης. Κατά τη δέσμηση του προσδέτη ενεργοποιούνται διάφορα καθοδικά (downstream) σηματοδοτικά μονοπάτια, τα οποία συνδέονται με τον πολλαπλασιασμό, την επιβίωση και την απόπτωση του κυττάρου^[44]. Πολυάριθμες in vitro και in vivo μελέτες ανέδειξαν τη λειτουργική σημασία της οικογένειας EGFR σε ένα ευρύ φάσμα καρκίνων καθώς συχνά υπερεκφράζονται σε καρκινικά κύτταρα και έτσι προάγουν τον πολλαπλασιασμό τους^[45]. Αυτό στάθηκε η αφορμή για την ανάπτυξη μορίων που στοχεύουν αυτούς τους υποδοχείς. Ένα από αυτά είναι το lapatinib που αναστέλλει την κινάση τυροσίνης των HER1 και HER2 και μπορεί να καταστέλλει την ανάπτυξη και την επιβίωση των καρκινικών κυττάρων. Η δυνατότητά ταυτόχρονης αναστολής των HER1 και HER2 του προσδίδουν ενισχυμένη αποτελεσματικότητα σε σχέση με μόρια που στοχεύουν μόνο ένα μέλος της οικογένειας HER^[46].

2 Υλικά και Μέθοδοι

2.1 Σύνολα δεδομένων CRISPR Knockout δοκιμασιών

Στην παρούσα πτυχιακή εργασία χρησιμοποιήθηκαν και αναλύθηκαν τρία σύνολα δεδομένων. Το πρώτο σύνολο δεδομένων είναι από επιτέλεση genome-wide CRISPR knockout δοκιμασιών σε κύτταρα ηπατοκυτταρικού καρκίνου σχεδιασμένα για τη σίγαση (knockout) του γονιδίου ATRX. Το δεύτερο σύνολο δεδομένων είναι από επιτέλεση genome-wide CRISPR knockout δοκιμασιών σε HeLa κύτταρα, με τις οποίες συγκρίνεται η επώαση με θεραπευτική αγωγή Olaparib με επώαση με DMSO. Το τρίτο σύνολο δεδομένων αντλήθηκε από τα αποτελέσματα genome-wide CRISPR knockout δοκιμασιών σε κύτταρα γαστρικού καρκίνου, με τις οποίες γίνεται σύγκριση της επώασης με Lapatinib ή με DMSO.

Τα παραπάνω σύνολα δεδομένων έχουν κατατεθεί στο Sequence Read Archive (SRA), το οποίο αποτελεί το μεγαλύτερο δημόσια διαθέσιμο αποθετήριο δεδομένων αλληλούχησης υψηλής απόδοσης^[47].

2.1.1 Το σύνολο δεδομένων ATRX Knockout

Τα συγκεκριμένα δεδομένα έχουν ανακτηθεί από την μελέτη με τίτλο: “Genome-Wide CRISPR-Cas9 Screen Reveals Selective Vulnerability of ATRX-Mutant Cancers to WEE1 Inhibition”^[37]. Χρησιμοποιήθηκε η κυτταρική σειρά PCL/PRF/5 ανθρώπινου ηπατοκυτταρικού καρκινώματος, η οποία συνθέτει το επιφανειακό αντιγόνο του ιού της ηπατίτιδας Β (HgsAB). Τα κύτταρα αυτά είχαν κατεργαστεί με σκοπό τη σίγαση του γονιδίου ATRX (ATRX knockout) και είχαν προσβληθεί με λεντιούς (lentiviruses) της ανθρώπινης βιβλιοθήκης GeCKO v2. Ως δείγμα αναφοράς (Control) συλλέχθηκαν τα δεδομένα αλληλούχησης από την ημέρα 0 της καλλιέργειας των κύτταρων, ενώ ως τελικό δείγμα αυτά από την ημέρα 21 (Treatment).

2.1.2 Το σύνολο δεδομένων επώασης με Olaparib

Τα δεδομένα έχουν ανακτηθεί από την μελέτη με τίτλο: “The chromatin remodeler ALC1 underlies resistance to PARP inhibitor treatment”^[42]. Σε αυτήν την έρευνα χρησιμοποιήθηκε η κυτταρική σειρά HeLa, γνωστή και ως η παλαιότερη, η πιο διαδεδομένη, “αθανατοποιημένη” ανθρώπινη κυτταρική σειρά^[48]. Προέρχεται από καρκινικά κύτταρα του τραχήλου της μήτρας, που ελήφθησαν το 1951 και πήρε το όνομά της από την Henrietta Lacks, μια 31χρονη Αφροαμερικανή μητέρα πέντε παιδιών. Η κυτταρική σειρά βρέθηκε να είναι εξαιρετικά ανθεκτική και παραγωγική, γεγονός που της επιτρέπει να χρησιμοποιείται εκτενώς σε επιστημονικές μελέτες. Στη συγκεκριμένη μελέτη μολύνθηκαν δύο βιολογικά αντίγραφα των HeLa κυττάρων αγρίου τύπου, με τη βιβλιοθήκη GeCKO v2 A και με την GeCKO v2 B (με χρήση λεντιών). Στη συνέχεια εφαρμόστηκε επώαση με τον αναστολέα PARP, Olaparib και έπειτα από 14 ημέρες συλλέχθηκε γονιδιωματικό DNA, το οποίο υποβλήθηκε σε αλληλούχηση (Treatment). Ως δείγμα ελέγχου (Control) χρησιμοποιήθηκε η επώαση με διμεθυλοσουλφοξείδιο (DMSO), έναν οργανικό διαλύτη με κρουπροστατευτική

δράση και έπειτα από 14 ημέρες συλλέχθηκε γονιδιωματικό DNA, το οποίο υποβλήθηκε σε αλληλούχηση.

2.1.3 Το σύνολο δεδομένων επώασης με *Lapatinib*

Τα δεδομένα έχουν ανακτηθεί από τη μελέτη με τίτλο: “A novel treatment strategy for lapatinib resistance in a subset of HER2-amplified gastric cancer”^[49]. Σε αυτήν την έρευνα δύο κυτταρικές σειρές γαστρικού καρκίνου, οι N87 και OE19, μολύνθηκαν από λεντιούς, που φέρουν τη βιβλιοθήκη GeCKO v2. Έπειτα, στα μετασχηματισμένα κύτταρα εφαρμόστηκε επώαση με *Lapatinib* (συνθήκη Treatment) και με DMSO (συνθήκη Control). Για κάθε κυτταρική σειρά πραγματοποιήθηκαν δύο ικές επιμολύνσεις σε δύο ξεχωριστά βιολογικά αντίγραφα. Μετά από 14 ημέρες συλλέχθηκε γονιδιωματικό DNA από όλα τα δείγματα και υποβλήθηκε σε αλληλούχηση. Τα δεδομένα που αντλήσαμε για την συγκεκριμένη ανάλυση αφορούν την OE19, η οποία αποτελεί κυτταρική σειρά αδενοκαρκινώματος του οισοφάγου.

2.2 Έλεγχος ποιότητας

Κατά την αλληλούχηση με Illumina, τεχνητές αλληλουχίες, όπως PCR υποκινητές (primers) και ακολουθίες ανταπτόρων (adapter sequences), συνδέονται σε επιλεγμένα βιολογικά θραύσματα, ώστε αυτά να ενισχυθούν και να συνδεθούν στο Illumina flow cell. Για να επιτευχθεί ακριβής ανάλυση είναι σημαντικό να εκτελέσουμε έλεγχο ποιότητας των ακατέργαστων δεδομένων, ελέγχοντας, μεταξύ άλλων, εάν αυτές οι τεχνητές αλληλουχίες έχουν αλληλουχηθεί μαζί με τα βιολογικά θραύσματα, ούτως ώστε να αφαιρεθούν.

Το εργαλείο που χρησιμοποιήθηκε για τον έλεγχο ποιότητας είναι το FastQC, το οποίο είναι κατάλληλο για δεδομένα αλληλουχιών υψηλής απόδοσης. Σε αυτό εισαγάγαμε τις αλληλουχίες, σε μορφότυπο αρχείου FASTQ, και μας παρείχε ένα σύνολο αναλύσεων για μια γρήγορη επισκόπηση των δεδομένων ώστε να εντοπίσουμε εάν και σε ποιους τομείς μπορεί να υπάρχουν προβλήματα. Στα αποτελέσματα του FastQC, ελέγχοντας εάν κάποιες αλληλουχίες εμφανίζονται σε μεγάλο ποσοστό (Overrepresented sequences) μπορούμε να εντοπίσουμε πιθανές τεχνητές αλληλουχίες.

2.3 Επεξεργασία δεδομένων

Απαραίτητο βήμα πριν την ανάλυση αποτελεί η επεξεργασία των δεδομένων αλληλουχιών, με σκοπό να κρατηθούν μόνο οι αλληλουχίες που αντιστοιχούν σε sgRNAs. Ως εργαλεία για την επεξεργασία χρησιμοποιήθηκαν οι εντολές *barcode splitter* και *fastxtrimmer* από το FASTX-Toolkit, καθώς και ο αλγόριθμος Cutadapt. Η εντολή *barcode splitter* δέχεται FASTA ή FASTQ αρχεία και τα χωρίζει σε επιμέρους μικρότερα αρχεία, βάσει αντιστοίχισης barcode και χρησιμοποιήθηκε, όπου χρειαζόταν, για τον αρχικό διαχωρισμό των αλληλουχιών. Έπειτα, χρησιμοποιήθηκε ο αλγόριθμος Cutadapt για το κόψιμο των

αλληλουχιών υποκινητή. Στις αλληλουχίες από το πρώτο σύνολο δεδομένων (ATRX knockout) εντοπίστηκε, βάσει του FastQC, η αλληλουχία υποκινητή “TCTTGTGGAAAGGACGAAACACCG” και αφαιρέθηκε μέσω του Cutadapt. Στις αλληλουχίες του δεύτερου και του τρίτου συνόλου δεδομένων (επώαση με Olaparib και με Lapatinib αντιστοίχα) ορίστηκε για περικοπή η αλληλουχία “CTTGTGGAAAGGACGAAACACCG”, με μερική αντιστοίχιση στην πρώτη βάση. Τέλος, για να διατηρηθούν μόνο τα πρώτα 20 νουκλεοτίδια στα οποία αντιστοιχούν τα sgRNAs των βιβλιοθηκών, χρησιμοποιήθηκε η εντολή *fastxtrimmer*. Έτσι, κρατηθήκαν σε κατεργασμένα FASTQ αρχεία μόνο οι επιθυμητές αλληλουχίες-στόχοι.

2.4 Εκτέλεση ανάλυσης MAGeCK

Χρησιμοποιήσαμε το εργαλείο MAGeCK για να αναλύσουμε τα δεδομένα από τα σύνολα δεδομένων: α) αποσιώπησης του ATRX, β) επώασης με Olaparib και γ) με Lapatinib.

Αρχικά, για κάθε σύνολο δεδομένων εκτελέσαμε την εντολή *mageck count* με την οποία αντιστοιχίσαμε τα διαβάσματα από τα FASTQ αρχεία με sgRNAs της ανθρώπινης βιβλιοθήκης GeCKO v2 παράγοντας έναν πίνακα μετρήσεων (count table) για κάθε σύνολο δεδομένων. Αυτοί οι πίνακες μετρήσεων περιέχουν τα sgRNAs που βρέθηκαν να αντιστοιχούν στην GeCKO v2, το όνομα του γονιδίου/miRNA στο οποίο στοχεύει το καθένα και το πλήθος διαβασμάτων ανά sgRNA που εντοπίζεται σε κάθε πειραματική συνθήκη.

Το επόμενο βήμα ήταν, με βάση τον παραπάνω πίνακα μετρήσεων, να συγκρίνουμε τις συνθήκες Control και Treatment, ανάλογα με τα δεδομένα του κάθε συνόλου:

- Για το σύνολο δεδομένων ATRX Knockout, συγκρίναμε τα δεδομένα από τις ημέρες από την αποσιώπηση 0 (δείγμα ελέγχου) και 21.
- Για το σύνολο δεδομένων Olaparib Treatment, συγκρίναμε τα δεδομένα από την επώαση με DMSO (ημέρα 14, δείγμα ελέγχου) και την επώαση με Olaparib (ημέρα 14).
- Για το σύνολο δεδομένων Lapatinib Treatment, συγκρίναμε τα δεδομένα από την επώαση με DMSO (ημέρα 14, δείγμα ελέγχου) και την επώαση με Lapatinib (ημέρα 14).

Κάναμε χρήση των αλγορίθμων MAGeCK RRA και MAGeCK MLE, ξεχωριστά. Με τους πίνακες μετρήσεων ως ορίσματα εκτελέσαμε αντιστοίχως τις εντολές *mageck test* (μέθοδος MAGeCK RRA) και *mageck mle* (μέθοδος MAGeCK MLE), προβάλλοντας με συγκεντρωτικό τρόπο (γονίδια και miRNAs μαζί) τα θετικά και τα αρνητικά επιλεγμένα γονίδια και miRNAs για το κάθε σύνολο δεδομένων.

Από τα αποτελέσματα της εντολής *mageck test*, για τους σκοπούς της συγκεκριμένης ανάλυσης, διατηρήσαμε το αρχείο “*gene_summary.txt*” και αντλήσαμε τις εξής πληροφορίες:

- Τα IDs των γονιδίων και των miRNAs.
- Τον αριθμό των sgRNAs που στοχεύουν στο κάθε γονίδιο ή miRNA.
- Τους βαθμούς “*neg|rank*” και “*pos|rank*” που προσδίδει η μέθοδος RRA στο κάθε γονίδιο και miRNA, οι οποίοι συμβολίζουν την κατάταξή τους στην αρνητική και θετική επιλογή αντίστοιχα.
- Τις τιμές σημαντικότητας “*neg|p-value*” και “*pos|p-value*” του κάθε γονιδίου και miRNA για την αρνητική και την θετική επιλογή αντίστοιχα.
- Τα False Discovery Rates “*neg|fdr*” και “*pos|fdr*” για το κάθε γονίδιο και miRNA στην αρνητική και στην θετική επιλογή αντίστοιχα.

Ομοίως, από τα αποτελέσματα της εντολής *mageck mle* διατηρήσαμε το αντίστοιχο αρχείο “*gene_summary.txt*” και αντλήσαμε τις εξής πληροφορίες:

- Τα IDs των γονιδίων και των miRNAs.
- Τον αριθμό των sgRNAs που στοχεύουν στο κάθε γονίδιο ή miRNA.
- Την βαθμολογία β score που προσδίδει η μέθοδος MLE, η οποία συνοψίζει τον βαθμό επιλογής για το κάθε γονίδιο και miRNA σε μία μεταβλητή (ανάλογα τον πειραματικό σχεδιασμό θετικές και αρνητικές τιμές β score αντιστοιχούν σε θετική/αρνητική επιλογή).
- Μια τιμή σημαντικότητας “*p-value*” για το κάθε γονίδιο και miRNA.
- Το False Discovery Rate (FDR) για το κάθε γονίδιο και miRNA.

2.5 Εκτέλεση ξεχωριστών αναλύσεων των miRNA και των γονιδιακών εγγραφών και σύγκριση αποτελεσμάτων

Επεξεργαστήκαμε τους αρχικούς πίνακες μετρήσεων από τα τρία σύνολα δεδομένων και χωρίσαμε τον καθένα σε δύο επιμέρους πίνακες, έναν για τις miRNA εγγραφές και έναν για τις γονιδιακές εγγραφές. Στη συνέχεια, πραγματοποιήσαμε ξεχωριστές αναλύσεις για τα miRNAs και για τα γονίδια εκτελώντας ξανά τις εντολές *mageck test* και *mageck mle* με ορίσματα τους νέους πίνακες, ξεχωριστά.

Συλλέξαμε τα αποτελέσματα (“*gene_summary.txt*” αρχεία) των συναρτήσεων RRA και MLE για miRNAs-γονίδια ξεχωριστά και για miRNAs-γονίδια συγκεντρωτικά και τα συγκρίναμε χρησιμοποιώντας την γλώσσα προγραμματισμού R. Παράξαμε γραφήματα διασποράς (scatterplots) για να συγκρίνουμε τους βαθμούς στην αρνητική κατάταξη (“*neg|rank*”) και στην θετική κατάταξη (“*pos|rank*”) των γονιδίων και των miRNAs, που προκύπτουν από τον συγκεντρωτικό RRA έλεγχο (γονίδια και miRNAs μαζί) με αυτούς που προκύπτουν από τον ξεχωριστό RRA έλεγχο (γονίδια και miRNAs ξεχωριστά). Παράλληλα,

παράξαμε επιπλέον γραφήματα διασποράς για να συγκρίνουμε τα β scores που έχουν αποδοθεί στο κάθε γονίδιο και miRNA από τον συγκεντρωτικό MLE έλεγχο με αυτά από τον ξεχωριστό MLE έλεγχο.

Έπειτα, εξετάσαμε εάν επηρεάζονται οι διαφορές-αποκλίσεις των βαθμών (rank difference) και των β score (beta score difference), αντίστοιχα, από τον μέσο αριθμό διαβασμάτων (average count reads) για κάθε γονίδιο και miRNA, παράγοντας επιπλέον γραφήματα διασποράς και ελέγχοντας τη γραμμικότητα (Linearity) των δεδομένων.

2.6 Σύγκριση των βαθμολογιών των intragenic miRNAs με αυτές των host γονιδίων τους

Χρησιμοποιώντας τα αποτελέσματα των συγκεντρωτικών αναλύσεων RRA και MLE σε συνδυασμό με βιβλιοθήκες του εργαστηρίου (in-house), οι οποίες περιγράφουν την αντιστοίχιση μεταξύ των intragenic miRNAs και των host γονιδίων, που τα φέρουν στην κωδική/μη-κωδική περιοχή τους, εκτελέσαμε επιμέρους ελέγχους για να εξετάσουμε εάν υπάρχει συσχέτιση ανάμεσα στις βαθμολογίες (ranks/ β scores) των intragenic miRNAs και στις βαθμολογίες (ranks/ β scores) των host γονιδίων τους. Παράξαμε γραφήματα διασποράς συγκρίνοντας για το κάθε ζεύγος intragenic miRNA και host γονιδίου την αντίστοιχη βαθμολογία που έχει αποδοθεί στο miRNA και αυτή στο γονίδιο. Πιο συγκεκριμένα, προβάλαμε με ξεχωριστά γραφήματα διασποράς: α) τις βαθμολογίες από τα ζεύγη όπου το miRNA μεταγράφεται με ίδια φορά μεταγραφής με το γονίδιο, β) τις βαθμολογίες από τα ζεύγη όπου το miRNA μεταγράφεται με αντίθετη φορά και γ) τις βαθμολογίες από όλα τα ζεύγη συνολικά ανεξάρτητα από τη φορά μεταγραφής. Επίσης, χρησιμοποιώντας τα αποτελέσματα των ξεχωριστών αναλύσεων RRA και MLE, εξετάσαμε εάν τα intragenic miRNAs που έχουν βρεθεί στατιστικώς σημαντικά και αρνητικά επιλεγμένα έχουν το αντίστοιχο host γονιδίο τους επίσης στατιστικώς σημαντικό και αρνητικά επιλεγμένο.

2.7 Εξειδικευμένη μελέτη χαρακτηριστικών των εγγραφών για τα miRNAs και τους στόχους τους

Έχοντας εκτελέσει τις παραπάνω συγκρίσεις-ελέγχους, εμβαθύναμε στις αλληλεπιδράσεις miRNAs και γονιδίων. Για αυτό το βήμα είναι απαραίτητο να υπάρχουν στα αποτελέσματά μας στατιστικώς σημαντικά miRNAs. Όμως, επειδή **(α)** το MAGeCK χρησιμοποιεί ένα αρκετά αυστηρό στατιστικό πλαίσιο για την αξιολόγηση της στατιστικής σημασίας και **(β)** πιθανόν η επιλογή των miRNAs και/ή των γονιδίων προς μία από τις δύο κατευθύνσεις δεν ευνοούνταν από την πειραματική-βιολογική συνθήκη, η εκτίμηση του FDR σε κάποιες περιπτώσεις κρίθηκε ότι ήταν αρκετά συντηρητική. Κατά περίπτωση, είτε στη μελέτη της μεθόδου RRA είτε της MLE, δεν προχωρήσαμε σε ανάλυση των miRNAs και των στόχων τους για το λόγο αυτό (Αποτελέσματα-Συζήτηση).

Για την εξειδικευμένη μελέτη των miRNA εγγραφών, αρχικά, χρησιμοποιώντας δεδομένα από τη miRBase, αντιστοιχίσαμε τις πρόδρομες μορφές των miRNAs με τις ώριμες μορφές τους για να απεικονιστούν και οι διαφορές στη σημαντικότητα των 5p/3p μορφών. Έπειτα, αντλήσαμε δεδομένα πρόβλεψης αλληλεπιδράσεων των miRNAs από τον αλγόριθμο DIANA-microT-CDS για να αντιστοιχίσουμε τα ώριμα miRNAs με τα γονίδια-στόχους τους (κατώφλι 0.7 στο score πρόβλεψης).

Στα RRA αποτελέσματα αξιοποιήσαμε miRNAs που εμφανίζονταν (α) υψηλά, (β) στη μέση και (γ) χαμηλά στην αρνητική κατάταξη (Negative Ranking) και δημιουργήσαμε θηκογράμματα (boxplots), που αναδεικνύουν την κατανομή των βαθμών θετικής κατάταξης (Positive Ranking) των γονιδίων-στόχων του κάθε miRNA ανάλογα με τη θέση που κατέχει αυτό στη διαβάθμιση (υψηλής, μεσαίας ή χαμηλής κατάταξης). Στα MLE αποτελέσματα διατηρήσαμε (α) ως ομάδα υψηλής σημαντικότητας τα miRNAs με αρνητικό β score και υψηλή στατιστική σημασία (FDR < 0.05 στο ATRX Knockout και FDR < 0.25 στο Lapatinib Treatment), (β) ως μεσαία ομάδα πήραμε αυτά στη μέση της κατάταξης και (γ) ως χαμηλή αυτά στο τέλος της κατάταξης. Έτσι, κατασκευάσαμε θηκογράμματα (boxplots), που αναδεικνύουν την κατανομή των β scores των γονιδίων-στόχων του κάθε miRNA ανάλογα με τη θέση που κατέχει αυτό στη διαβάθμιση (υψηλής, μεσαίας ή χαμηλής κατάταξης). Επιπλέον, πράξαμε γραφήματα διασποράς δείχνοντας για το κάθε miRNA, ανάλογα με τη θέση του στην διαβάθμιση, τον υψηλότερο σε βαθμολογία γονιδιακό στόχο του και την βαθμολογία που κατέχει αυτός. Στην RRA ανάλυση ως υψηλότερος γονιδιακός στόχος του κάθε miRNA τέθηκε αυτός με την υψηλότερη θέση στην θετική κατάταξη (Positive Ranking), ενώ στην MLE αυτός με το μεγαλύτερο θετικό β score.

Τέλος, μέσω χαρτών θερμότητας (heat-maps), απεικονίσαμε ζεύγη miRNAs και στόχων που εμφάνιζαν σημαντική επιλογή, ως πιθανά ζεύγη που συσχετίζονται με τους υπό μελέτη φαινότυπους.

2.8 Εκτέλεση αναλύσεων εμπλουτισμού μονοπατιών με το MAGeCK Flute και το miRPath και αντιστοίχιση αποτελεσμάτων:

Στα αποτελέσματα των αλγορίθμων RRA και MLE για τις γονιδιακές εγγραφές από το κάθε σύνολο δεδομένων (ATRX Knockout και επώαση με Lapatinib), εκτελέσαμε ανάλυση εμπλουτισμού μονοπατιών χρησιμοποιώντας το πακέτο MAGeCK Flute.

Για τα αποτελέσματα RRA:

α) εκτελέσαμε τη συνάρτηση *ReadRRA()* του πακέτου MAGeCK Flute, η οποία διαβάζει το αρχείο "gene_summary" από τα RRA αποτελέσματα και επιστρέφει ένα LFC (Log Fold Change) score για κάθε γονίδιο

β) φιλτράραμε τα αποτελέσματα, κρατώντας μόνο τα θετικά επιλεγμένα στατιστικώς σημαντικά (LFC > 0, pos|pvalue < 0.05) γονίδια

γ) εκτελέσαμε την συνάρτηση *EnrichAnalyzer()* του πακέτου MAGeCK Flute για τα επιλεγμένα γονίδια, επιλέγοντας την πηγή μονοπατιών KEGG.

Για τα αποτελέσματα MLE:

α) εκτελέσαμε την συνάρτηση *ReadBeta()* του πακέτου MAGeCK Flute, η οποία διαβάζει το αρχείο “gene_summary” από τα MLE αποτελέσματα και επιστρέφει το β score (treatment) για κάθε γονίδιο

β) εκτελέσαμε τη συνάρτηση *IncorporateDermap()* (του MAGeCK Flute), η οποία επισημαίνει σε κάθε γονίδιο ένα control β score με βάση την κυτταρική σειρά

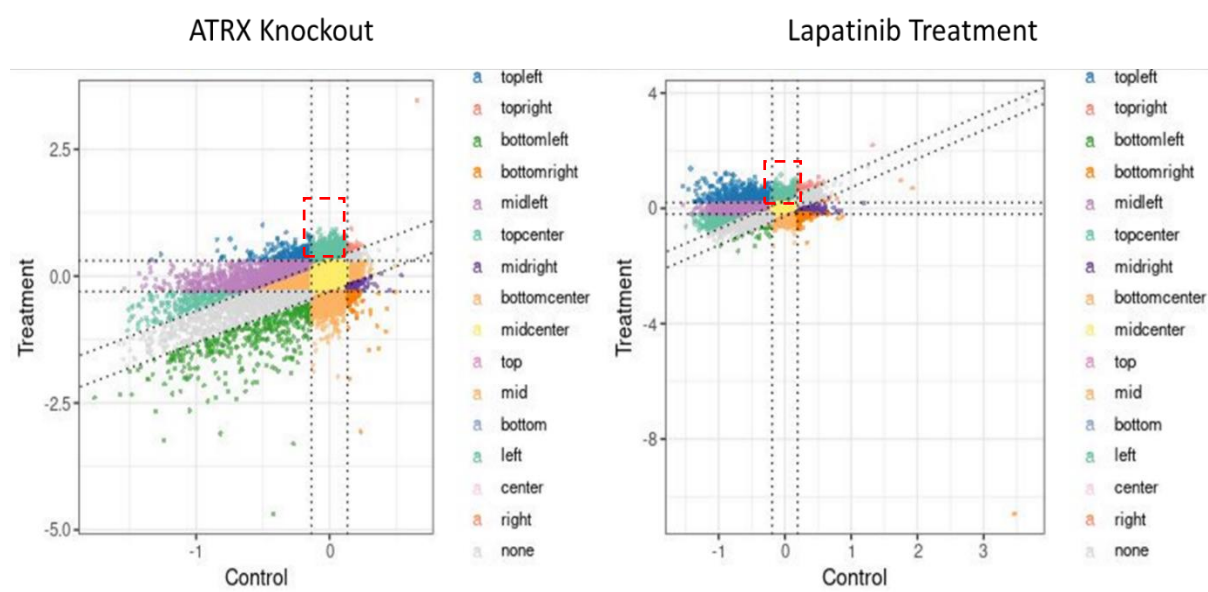
γ) εφαρμόσαμε τη συνάρτηση *NormalizeBeta()* (του MAGeCK Flute) για την κανονικοποίηση των β scores (control και treatment)

δ) με τη συνάρτηση *CutoffCalling()* υπολογίσαμε όρια αποκοπής και χρησιμοποιώντας την συνάρτηση *ScatterView()* (του MAGeCK Flute), δημιουργήσαμε γράφημα διασποράς με τα control και τα treatment β scores του κάθε γονιδίου (**Εικόνα 4**)

ε) διατηρήσαμε μόνο τα γονίδια τα οποία έχουν υψηλό θετικό treatment β score, αλλά χαμηλό θετικό/αρνητικό control β score

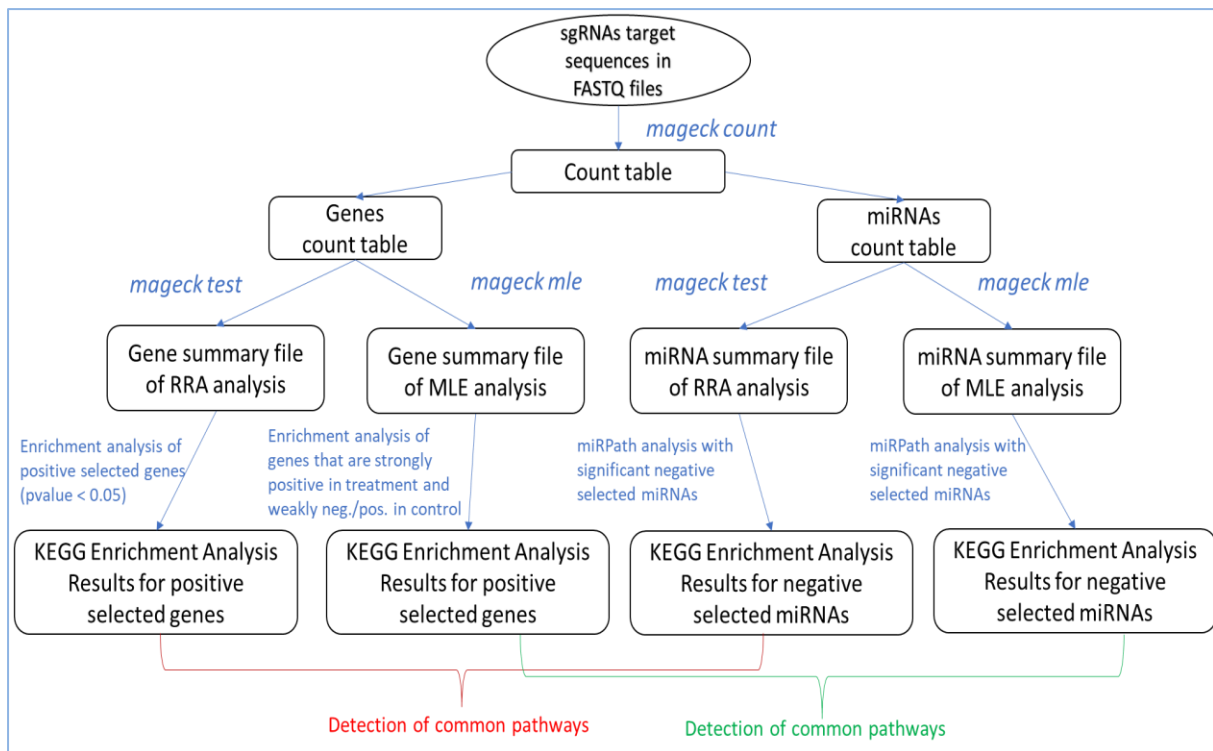
στ) εκτελέσαμε την συνάρτηση *EnrichAnalyzer()* για τα επιλεγμένα γονίδια, επιλέγοντας την πηγή μονοπατιών KEGG.

Από τα αποτελέσματα του RRA και του MLE για τις miRNA εγγραφές διατηρήσαμε, από το κάθε σύνολο δεδομένων (ATRX Knockout και επώαση με Lapatinib), τα στατιστικώς σημαντικά αρνητικά επιλεγμένα miRNAs. Έπειτα, εκτελέσαμε ανάλυση εμπλουτισμού για KEGG μονοπάτια στα επιλεγμένα miRNAs, χρησιμοποιώντας το miRPath (με κατώφλι για microT-CDS 0.7 και p-value όριο (FDR) < 0.05). Η συνολική ροή εργασίας που ακολουθήθηκε για την ταυτόχρονη ανάλυση συσχέτισης miRNAs και γονιδίων με συγκεκριμένους φαινότυπους υπό μελέτη παρουσιάζεται στην **Εικόνα 5**.



Εικόνα 4: Γραφήματα διασποράς για τον διαχωρισμό των γονιδίων από τις αναλύσεις MLE. Το γράφημα A αντιστοιχίζει τα Control με τα Treatment β scores των γονιδίων από τα δεδομένα ATRX

Knockout μετά την ανάλυση MLE, ενώ το γράφημα B αυτά από τα δεδομένα Laratinib Treatment μετά την ανάλυση MLE. Οι διακεκομμένες γραμμές συμβολίζουν τα όρια αποκοπής που έχουν υπολογιστεί για τα Treatment β scores (οριζόντια), τα Control β scores (κάθετα) και για τις διαφορές τους (Treatment – Control, διαγώνια). Βάσει αυτών διαχωρίζονται (και χρωματικά) ομάδες γονιδίων. Από τις ομάδες γονιδίων επιλέγονται προς ανάλυση εμπλουτισμού μονοπατιών αυτές που σημαίνονται με κόκκινες διακεκομμένες γραμμές, δηλαδή γονίδια με υψηλό θετικό Treatment β score, αλλά ήπιο (θετικό ή αρνητικό) Control β score. Πρόκειται για γονίδια των οποίων η απώλεια λειτουργείας συσχετίζεται με επιβίωση-ανθεκτικότητα στις επιδράσεις του φαινοτύπου Treatment (περισσότερες λεπτομέρειες στην εργασία αναφοράς^[9]).

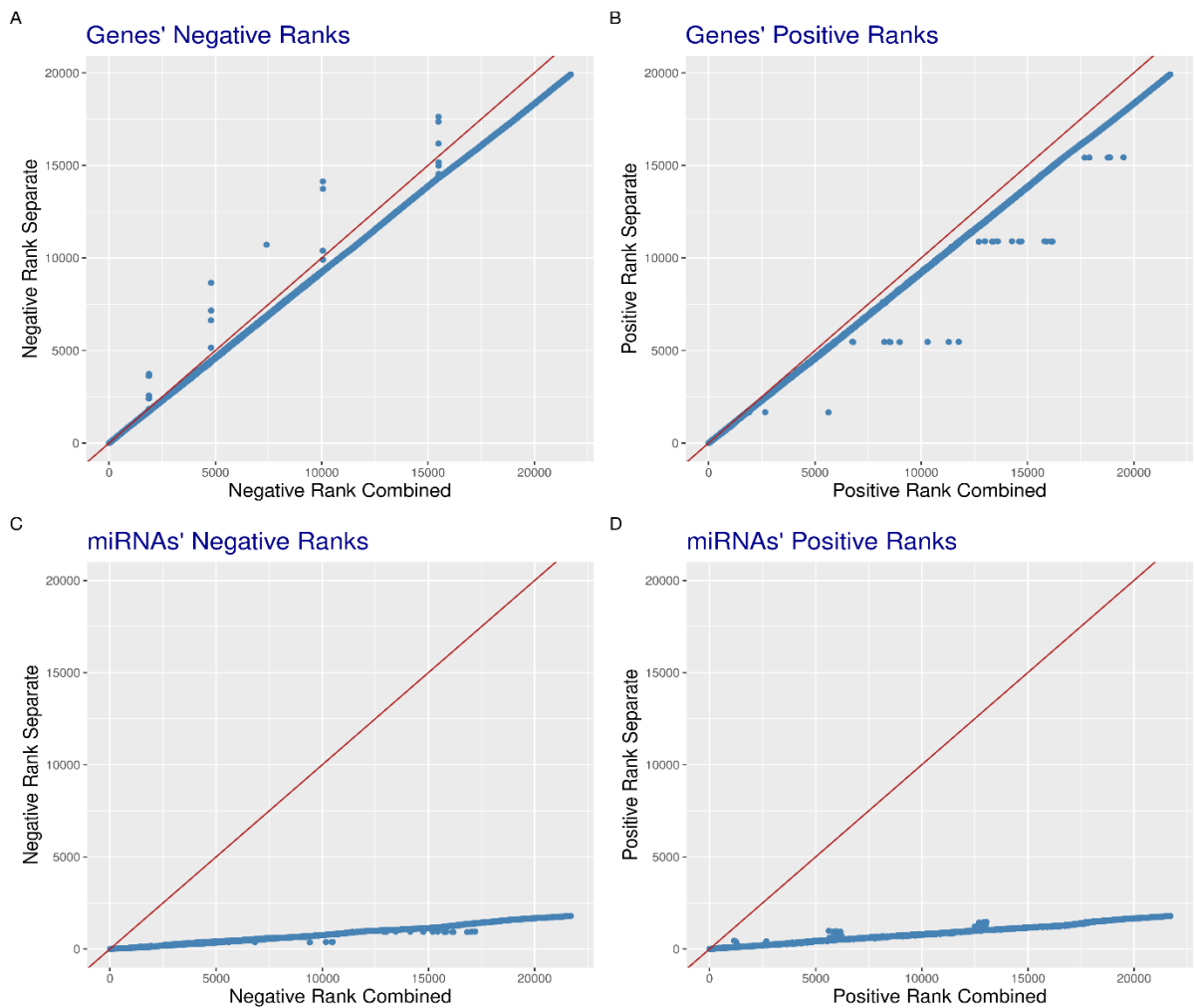


Εικόνα 5: Σχηματική αναπαράσταση της διαδικασίας της ξεχωριστής ανάλυσης των γονιδίων και των miRNAs με τις μεθόδους RRA και MLE και της ανάλυσης εμπλουτισμού των αποτελεσμάτων. Τα εισαγόμενα δεδομένα (input data) στην ανάλυση είναι τα επεξεργασμένα αρχεία FASTQ που περιέχουν μόνο τις αλληλουχίες στις οποίες στοχεύουν τα sgRNAs. Για την ανάλυση των δεδομένων χρησιμοποιούνται οι συναρτήσεις MAGeCK RRA και MAGeCK MLE μέσω των εντολών mageck test και mageck mle, αντίστοιχα. Έπειτα, για τις αναλύσεις εμπλουτισμού KEGG μονοπατιών αξιοποιείται το πακέτο MAGeCKFlute και το λογισμικό miRPath.

3 Αποτελέσματα - Συζήτηση

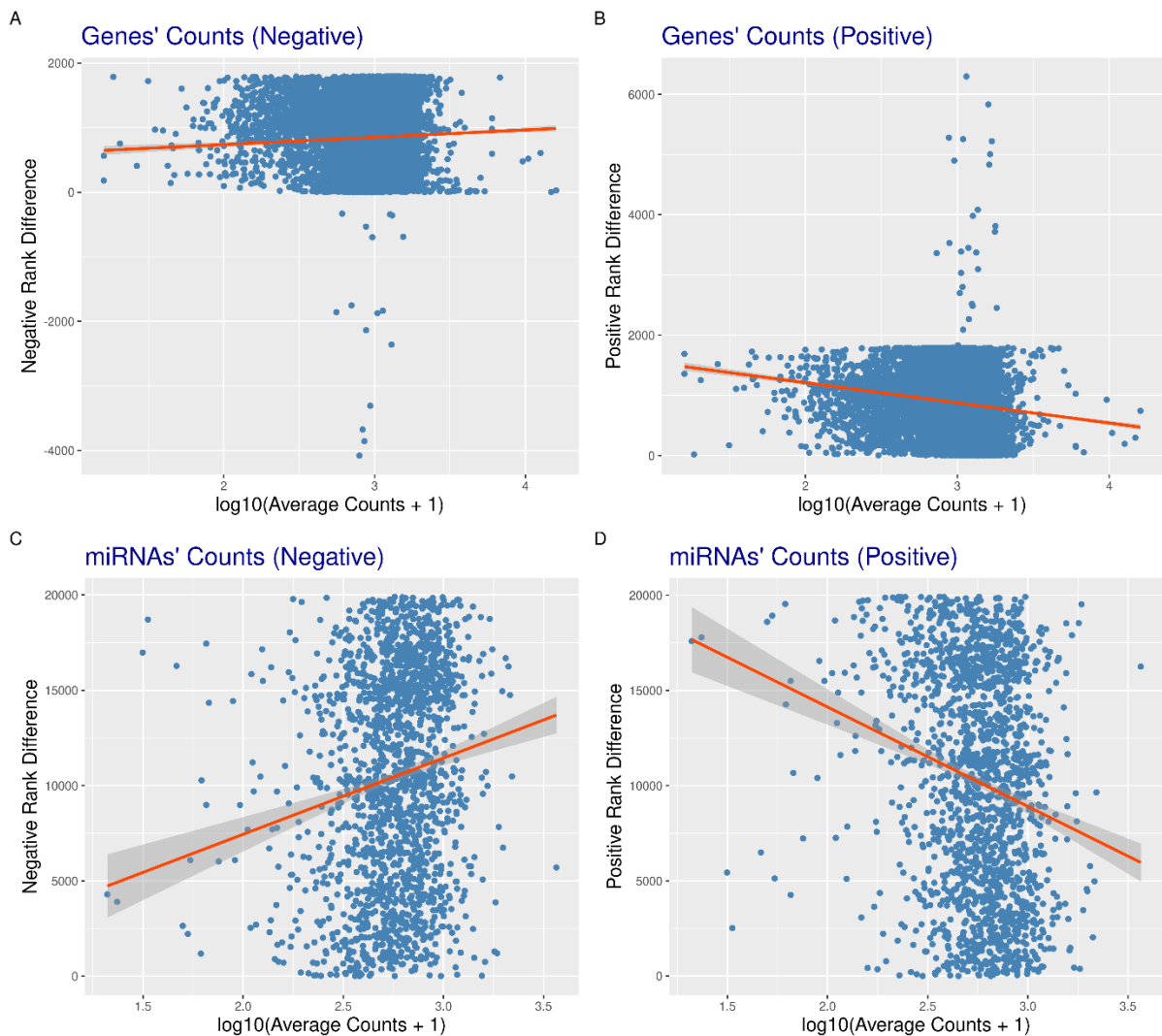
3.1 Σύγκριση αναλύσεων miRNAs και γονιδίων από κοινού και χωριστά

Οι συγκρίσεις των αποτελεσμάτων των ξεχωριστών και των συγκεντρωτικών αναλύσεων πραγματοποιήθηκαν για να ελέγξουμε κατά πόσο επηρεάζονται οι βαθμολογίες των γονιδίων και των miRNAs από τον διαχωρισμό των εγγραφών. Στα αποτελέσματα του αλγορίθμου MAGeCK RRA, οι βαθμοί στην αρνητική και στη θετική κατάταξη των γονιδίων και των miRNAs εξαρτώνται από το πλήθος των εγγραφών. Διατηρώντας μόνο τις γονιδιακές εγγραφές, δεν αποδίδονται (άρα «χάνονται») οι βαθμοί κατάταξης των miRNAs με αποτέλεσμα οι γονιδιακές εγγραφές να ανεβαίνουν θέση στην κατάταξη (δηλαδή να μειώνονται οι βαθμοί κατάταξής τους). Το ίδιο ισχύει και αντίστροφα για την απομόνωση μόνο των miRNA εγγραφών. Και για τα τρία σύνολα δεδομένων (ATRX Knockout, Olaparib Treatment, Lapatinib Treatment) δεν παρατηρήθηκαν σημαντικές ανομοιομορφίες στις αποκλίσεις των βαθμών κατάταξης, πέρα από τις αναμενόμενες αποκλίσεις λόγω της αλλαγής του πλήθους των εγγραφών. Ενδεικτικά παρουσιάζονται τα αποτελέσματα από την σύγκριση στα δεδομένα ATRX Knockout (**Εικόνα 6**). Μικρές αποκλίσεις εμφανίζονται σε περιορισμένο αριθμό γονιδίων και miRNAs (οι περιπτώσεις που παρουσιάζονται απομακρυσμένες από την κύρια τάση των βαθμολογιών σε κάθε περίπτωση στην Εικόνα 6). Αυτές αποδίδονται στο γεγονός ότι κατά το MAGeCK RRA οι μέσες τιμές και η διασπορά όλων των διαθέσιμων sgRNAs αξιοποιείται για τη μοντελοποίηση της διασποράς.



Εικόνα 6: Σύγκριση negative/positive κατάταξης (ranking) στην ξεχωριστή και στην συγκεντρωτική ανάλυση MAGeCK RRA. Στα A και B scatterplots αναδεικνύεται η απόκλιση που υπάρχει, αντίστοιχα, ανάμεσα στα negative και στα positive ranks γονιδίων, που προκύπτουν από την ξεχωριστή ανάλυση (μόνο με γονιδιακές εγγραφές – separate ranks) και αυτών που προκύπτουν από τη συγκεντρωτική ανάλυση (γονιδιακές και miRNAs εγγραφές μαζί – combined ranks). Στα C και D scatterplots αναδεικνύεται η απόκλιση που υπάρχει ανάμεσα στα ranks των miRNAs, που προκύπτουν από την ανάλυση μόνο με miRNA εγγραφές (separate ranks) και αυτών που προκύπτουν από τη συγκεντρωτική ανάλυση, με γονιδιακές και miRNA εγγραφές (combined ranks). Η κόκκινη γραμμή στα γραφήματα χρησιμοποιείται ως απεικονιστικό βοήθημα για να αναδεικνύεται η απόκλιση των ranks.

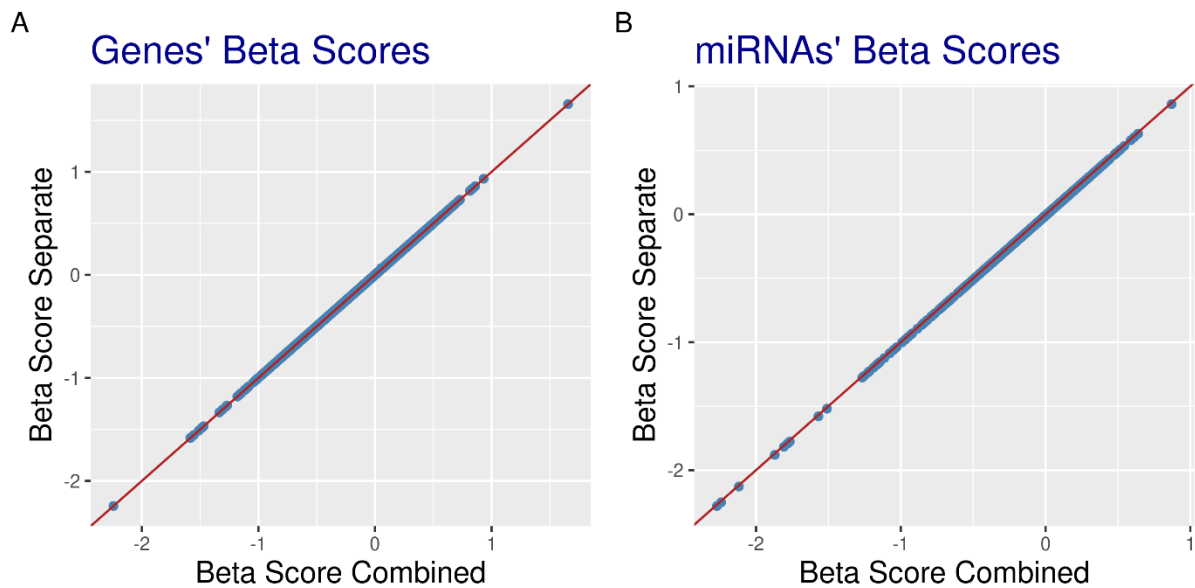
Ακολουθώντας, ελέγξαμε κατά πόσο η απόκλιση στους βαθμούς εμφανίζει γραμμικότητα σε σχέση με τον αριθμό των διαβασμάτων των sgRNAs. Στα παρακάτω γραφήματα απεικονίζεται η σχέση των αποκλίσεων στους βαθμούς των ξεχωριστών και των συγκεντρωτικών αναλύσεων με το πλήθος διαβασμάτων του κάθε γονιδίου και miRNA. Ενσωματώθηκε η γραμμή από τη μέθοδο γραμμικής παλινδρόμησης (Linear Regression) ως παράδειγμα γραμμικότητας. Και στα τρία σύνολα δεδομένων δεν εντοπίστηκε σημαντική γραμμική συσχέτιση. Ενδεικτικά παρουσιάζονται τα γραφήματα από το σύνολο δεδομένων ATRX Knockout (Εικόνα 7).



Εικόνα 7: Έλεγχος συσχέτισης της απόκλισης των βαθμών με το πλήθος των εγγραφών από τα αποτελέσματα του MAGeCK RRA. Στα γραφήματα A και C απεικονίζεται η αντιστοιχία της διαφοράς των αρνητικών βαθμών (*Negative Rank Difference*) και του μέσου αριθμού διαβασμάτων για το κάθε γονίδιο ή miRNA (*Average Counts*). Στα γραφήματα B και D απεικονίζεται η αντιστοιχία της διαφοράς των θετικών βαθμών (*Positive Rank Difference*) και του μέσου αριθμού διαβασμάτων για το κάθε γονίδιο ή miRNA (*Average Counts*). Σε κάθε περίπτωση, ως *Rank Difference* έχει οριστεί η διαφορά ανάμεσα στον βαθμό του κάθε γονιδίου ή miRNA στη συγκεντρωτική ανάλυση και σε αυτόν στη ξεχωριστή ανάλυση (*Rank Combined – Rank Separate*). Ως *Average counts* έχει οριστεί ο μέσος αριθμός διαβασμάτων του κάθε γονιδίου ή miRNA από τις συνθήκες του κάθε πειράματος. Επειδή το εύρος των διαβασμάτων ήταν πολύ μεγάλο απεικονίζονται μετασχηματισμένα στη λογαριθμική κλίμακα με βάση το 10.

Αντίστοιχα, στα αποτελέσματα του MAGeCK MLE απεικονίστηκαν οι αποκλίσεις στα β scores των γονιδίων και των miRNAs από την ξεχωριστή και την συγκεντρωτική ανάλυση. Η απόδοση των β scores από τον αλγόριθμο MLE είναι ειδική για κάθε γονίδιο και miRNA και δεν εξαρτάται από το συνολικό πλήθος των εγγραφών, αλλά από το πλήθος των διαβασμάτων που αναλογούν σε sgRNAs ανά τις συνθήκες των πειραμάτων. Μεταξύ των ξεχωριστών και των συγκεντρωτικών αναλύσεων MLE οι αποκλίσεις στα β scores ήταν

μηδενικές και στα τρία σύνολα δεδομένων, ενδεικτικά παρουσιάζονται αυτά που αποδοθήκαν στα δεδομένα ATRX Knockout (**Εικόνα 8**).

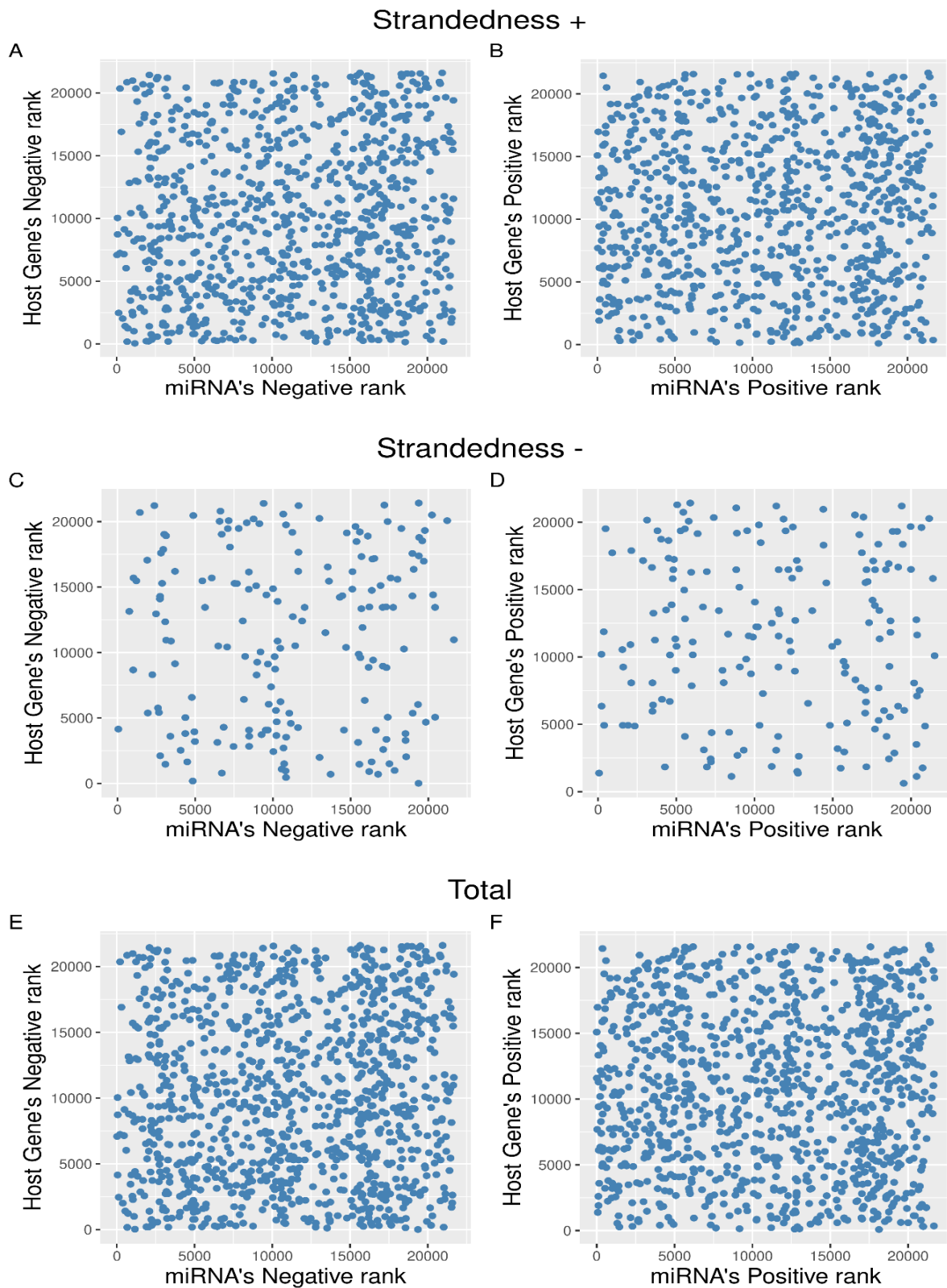


Εικόνα 8: Σύγκριση των β scores στην ξεχωριστή και στην συγκεντρωτική ανάλυση MAGeCK MLE. Στο scatterplot A αναδεικνύεται η απόκλιση που υπάρχει ανάμεσα στα β scores των γονιδίων που προκύπτουν από την ξεχωριστή ανάλυση (μόνο με γονιδιακές εγγραφές – Beta Score Separate) και αυτών που προκύπτουν από τη συγκεντρωτική ανάλυση (γονιδιακές και miRNAs εγγραφές – Beta Score combined). Στο scatterplot B αναδεικνύεται η απόκλιση που υπάρχει ανάμεσα στα β scores των miRNA, που προκύπτουν από την ανάλυση μόνο με miRNA εγγραφές (Beta Score Separate) και αυτών που προκύπτουν από τη συγκεντρωτική ανάλυση, με γονιδιακές και miRNA εγγραφές (Beta Score Combined). Η κόκκινη γραμμή στα γραφήματα χρησιμοποιείται ως απεικονιστικό βοήθημα για να αναδεικνύεται η απόκλιση των β scores.

3.2 Έλεγχος συσχέτισης βαθμολογιών των intragenic miRNAs και των γονιδίων τους

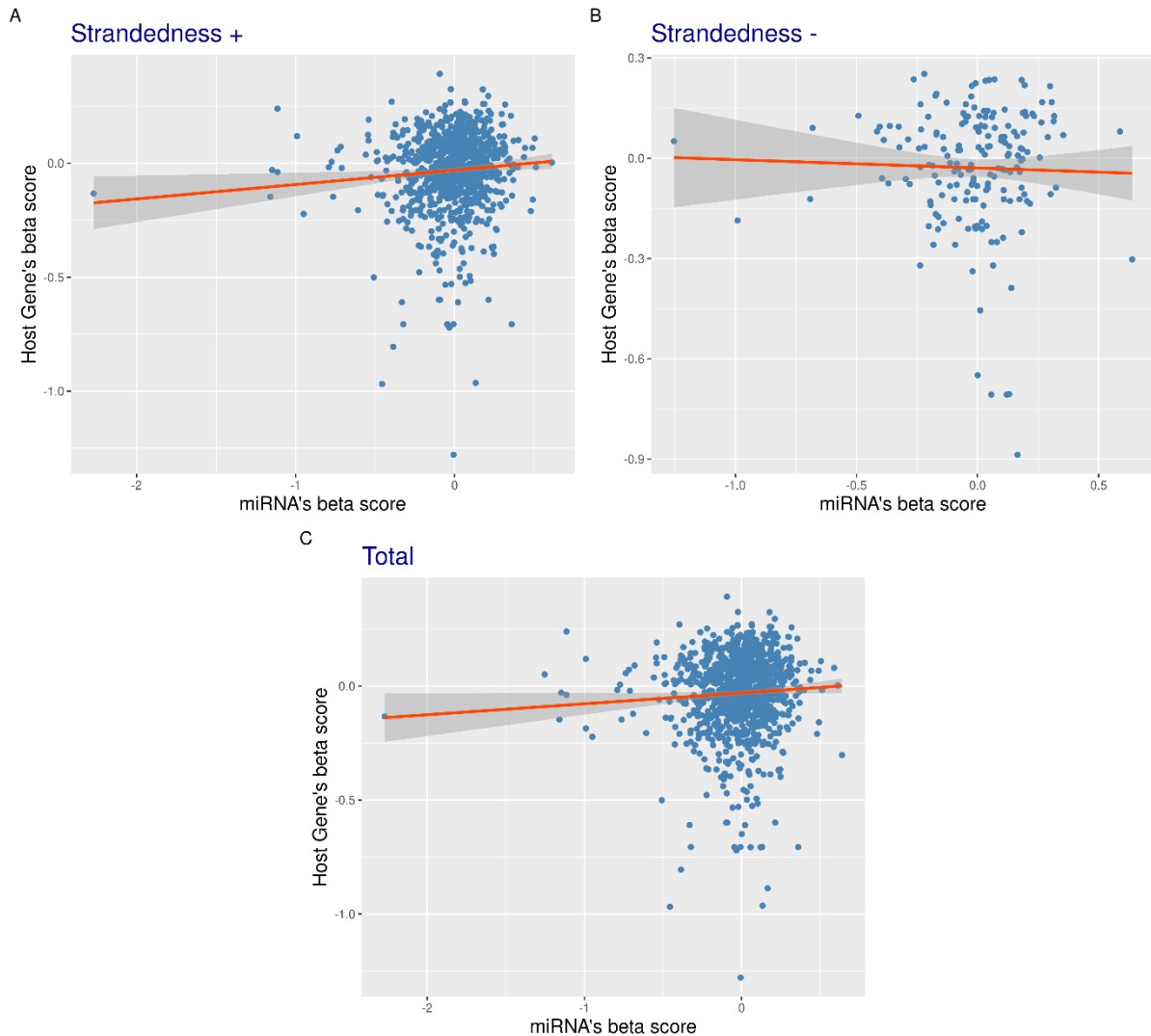
Τα intragenic miRNAs και τα host γονιδιά τους πολλές φορές συνεκφράζονται, γεγονός που θα μπορούσε να επηρεάζει την ποσοτικοποίηση των sgRNAs που στοχεύουν σε αυτά. Σκοπός μας ήταν να ελέγξουμε εάν στα αποτελέσματά μας εντοπίζονται τάσεις επιρροής ανάμεσα στις βαθμολογίες που έχουν αποδοθεί από τους αλγορίθμους RRA και MLE για τα intragenic miRNAs και για τα αντίστοιχα host γονιδιά τους. Έτσι, χρησιμοποιήσαμε τα γραφήματα που απεικονίζουν τα ranks και τα β scores αυτών κατ' αντιστοιχία. Στα συγκεκριμένα σύνολα δεδομένων, δεν εντοπίστηκε κάποια συνολική τάση συσχέτισης των βαθμών κατάταξης (ranks) ή των β scores των intragenic miRNAs με αυτών των host γονιδίων τους. Ενδεικτικά αυτό διακρίνεται στα αποτελέσματα από το σύνολο δεδομένων ATRX Knockout (**Εικόνα 9**, **Εικόνα 10**). Επιπλέον, ελέγξαμε και συγκεκριμένα εάν τα στατιστικώς σημαντικά miRNAs έχουν τα αντίστοιχα host γονιδιά τους ομοίως σημαντικά. Σε κανένα από

τα τρία σύνολα δεδομένων δεν αναδείχτηκε κάποια τέτοιου χαρακτήρα σχέση, που θα μπορούσε να επηρεάσει τα αποτελέσματά μας.



Εικόνα 9: Έλεγχος συσχέτισης της κατάταξης (θετικής/αρνητικής) των intragenic miRNAs και των αντίστοιχων host γονιδίων τους στα αποτελέσματα του MAGeCK RRA. Στα γραφήματα διασποράς

A και B απεικονίζεται η αντιστοιχία των βαθμών αρνητικής και θετικής κατάταξης, αντίστοιχα, των intragenic miRNAs και των host γονιδίων τους που βρίσκονται στην ίδια φορά μεταγραφής (Strandedness +). Στα γραφήματα C και D απεικονίζεται η αντιστοιχία των βαθμών αρνητικής και θετικής κατάταξης αυτών που εντοπίζονται σε αντίθετη φορά μεταγραφής (Strandedness -). Στα γραφήματα E και F φαίνονται συνολικά οι αντιστοιχίες βαθμών αρνητικής και θετικής κατάταξης των intragenic miRNAs και των host γονιδίων τους με ίδια ή αντίθετη κατεύθυνση.



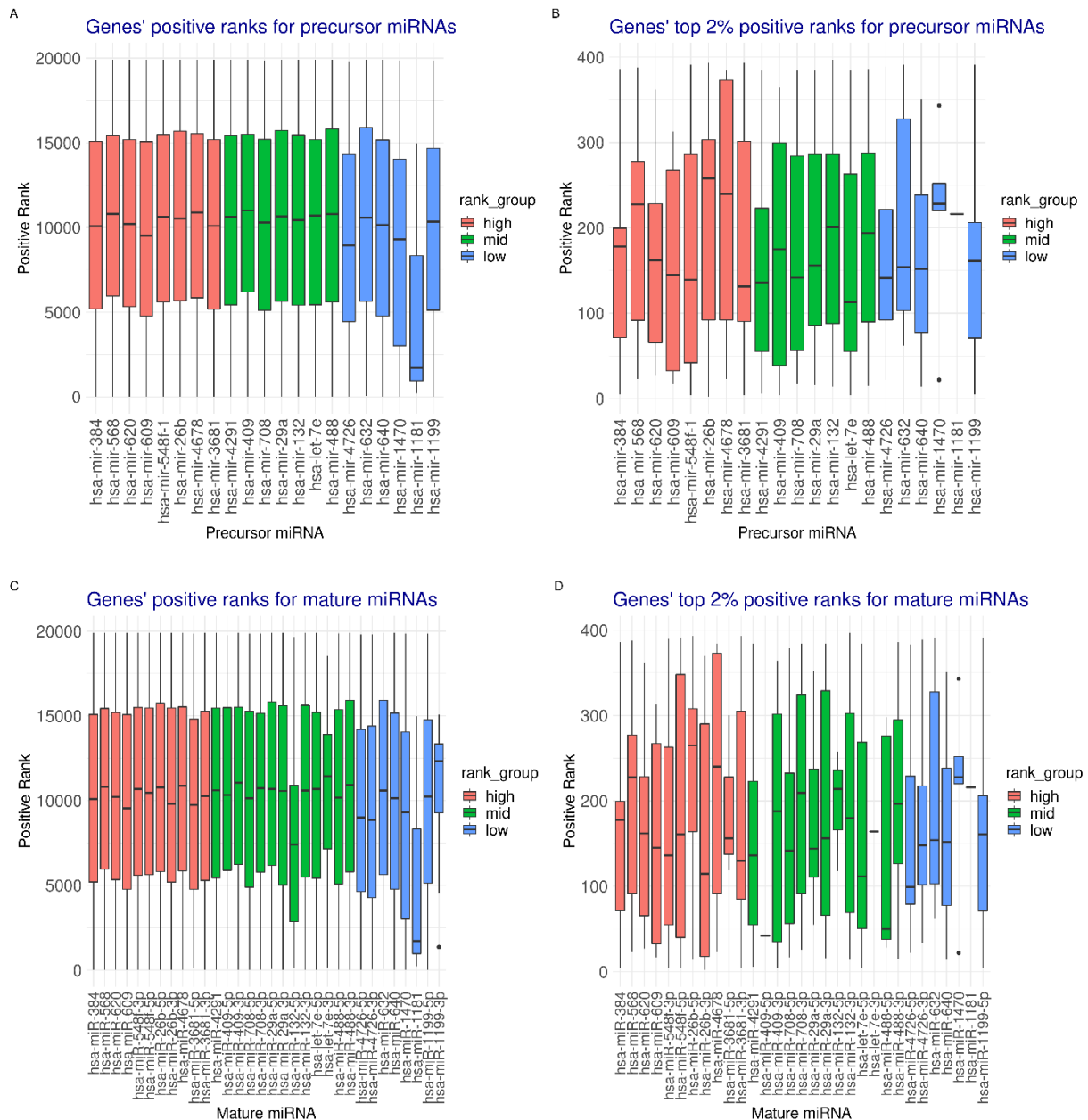
Εικόνα 10: Έλεγχος συσχέτισης μεταξύ των β scores των intragenic miRNAs και των αντίστοιχων host γονιδίων τους στα αποτελέσματα του MAGeCK MLE και εφαρμογή γραμμικής παλινδρόμησης. Στο γράφημα A ελέγχεται η συσχέτιση αναμεσα στα β scores των intragenic miRNAs και των host γονιδίων τους που εντοπίζονται στην ίδια φορά μεταγραφής (Strandedness +), ενώ στο γράφημα B ελέγχεται η συσχέτιση αναμεσα στα β scores αυτών που βρίσκονται σε αντίθετη φορά μεταγραφής (Strandedness -). Στο γράφημα C ελέγχεται η συνολική συσχέτιση των β scores των intragenic miRNAs και των host γονιδίων τους με ίδια ή αντίθετη φορά μεταγραφής. Έχουν ενσωματωθεί γραμμές που προκύπτουν από την εφαρμογή γραμμικής παλινδρόμησης.

3.3 Κατανομή των βαθμολογιών των γονιδίων-στόχων ανά miRNA

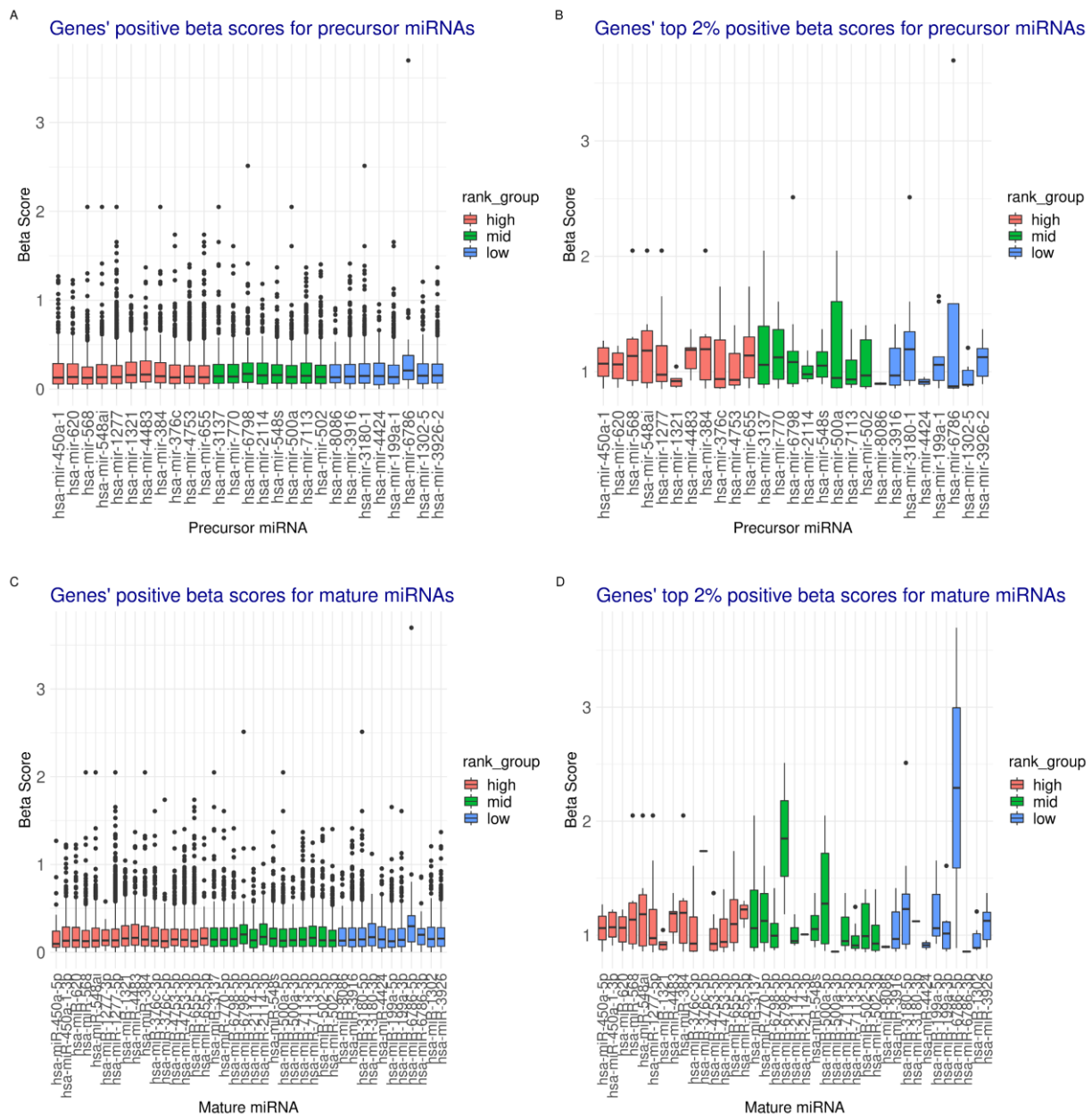
Τόσο στα αποτελέσματα του αλγορίθμου RRA όσο και στου MLE, διατηρήσαμε δείγματα miRNAs που εμφανίζουν υψηλή, μεσαία και χαμηλή διαβάθμιση στις βαθμολογίες. Ο αρχικός στόχος ήταν να ελέγξουμε εάν υπάρχει κάποια συσχέτιση στο εύρος ή/και στις διαμέσους των βαθμολογιών των γονιδίων-στόχων του κάθε miRNA ανάλογα με την κατάταξή τους. Στα αποτελέσματα μας δεν παρατηρήθηκε κάποια τάση διαφοροποίησης των βαθμολογιών των γονιδίων-στόχων ανάλογη με τη κατάταξη των miRNAs. Αυτό προκύπτει διότι το κάθε miRNA έχει πολλούς στόχους και συμβαίνουν αρκετές επικαλύψεις των στόχων των διαφορετικών miRNAs ακόμη και εάν αυτά βρίσκονται σε διαφορετική θέση κατάταξης. Η διάκριση που εκπορεύεται από τα γραφήματα αφορά τις ώριμες μορφές 5p και 3p των miRNAs. Στα αρχικά αποτελέσματα των αλγορίθμων RRA και MLE καταγράφονταν μόνο οι πρόδρομες μορφές των miRNAs, ενώ με την αντιστοίχιση που εκτελέσαμε απεικονίζεται η διαφοροποίηση στην κατανομή των γονιδιακών στόχων μεταξύ των ώριμων μορφών (5p/3p) των miRNAs. Ενδεικτικό παράδειγμα αποτελούν τα γραφήματα από τα δεδομένα Lapatinib Treatment (**Εικόνα 11, Εικόνα 12**).

Επιπλέον, τα παραπάνω ώριμα miRNAs αντιστοιχήθηκαν με τον υψηλότερο σε κατάταξη (positive rank/ β score) θετικά επιλεγμένο γονιδιακό τους στόχο. Με αυτόν τον τρόπο απεικονίστηκαν οι διαβαθμίσεις στην κατανομή των πιο σημαντικών γονιδιακών στόχων για το κάθε miRNA ανάλογα με την κατάταξή του (**Εικόνα 13, Εικόνα 14**). Στα γραφήματα μπορούμε να δούμε ποιος είναι ο πιο σημαντικός γονιδιακός στόχος για το κάθε miRNA καθώς και την βαθμολογία (θετικής επιλογής) του. Σε κανένα από τα δύο σύνολα δεδομένων δεν παρατηρήθηκε διαβάθμιση των στόχων σχετική με την κατάταξη των miRNAs.

Συμπεραίνουμε ότι, αξιοποιώντας προβλέψεις στόχων των miRNAs, δεν προκύπτει μια άμεσα μεταφράσιμη σχέση μεταξύ των miRNAs που εμφανίζουν σημαντική θετική/αρνητική επιλογή και αθροιστικά των στόχων τους, ή του σημαντικότερου στόχου βάσει βαθμολογίας. Είναι πιθανό τα ευρήματα να ήταν διαφορετικά, αν οι δοκιμασίες CRISPR Knockout συνδυάζονταν με πειράματα για την ανάδειξη στόχων των miRNAs (π.χ. πειράματα HITS-CLIP) εξειδικευμένα στην πειραματική συνθήκη υπό μελέτη.

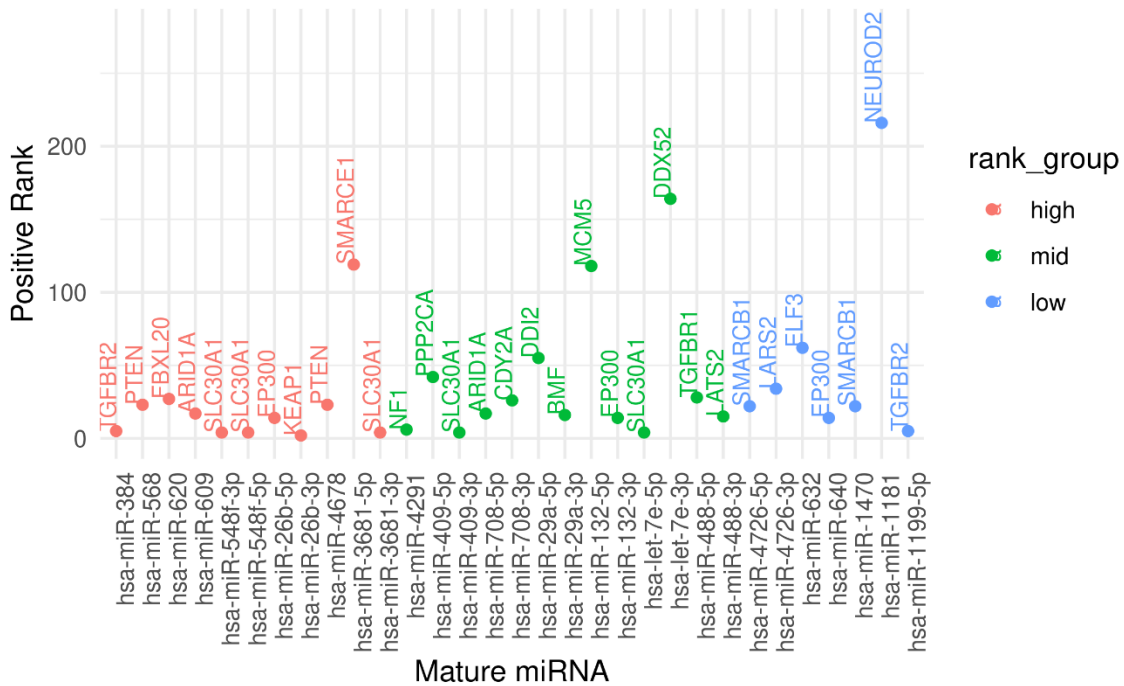


Εικόνα 11: Κατανομές των βαθμών θετικής κατάταξης των γονιδίων-στόχων του κάθε miRNA ανάλογα με τη θέση που διαθέτει στην κατάταξη. Το γράφημα A περιλαμβάνει τις κατανομές των βαθμών θετικής κατάταξης (positive ranks) των γονιδίων-στόχων αντιστοιχώντας τους στις πρόδρομες μορφές των miRNAs που έχουν απομονωθεί από κάθε κατάταξη, ενώ το γράφημα B περιλαμβάνει τις κατανομές μόνο των 2% υψηλότερων σε κατάταξη positive ranks. Στα γραφήματα C και D γίνεται η μετάβαση από τις πρόδρομες μορφές των miRNAs στις ώριμες και παρουσιάζονται οι αντίστοιχες κατανομές. Οι χρωματισμοί των miRNAs έχουν γίνει με βάση την θέση τους στην αρνητική κατάταξη (negative ranking): τα ροζ είναι στην υψηλότερη θέση (high), τα πράσινα είναι στη μέση (mid) και τα γαλάζια είναι στην κατώτερη θέση (low).

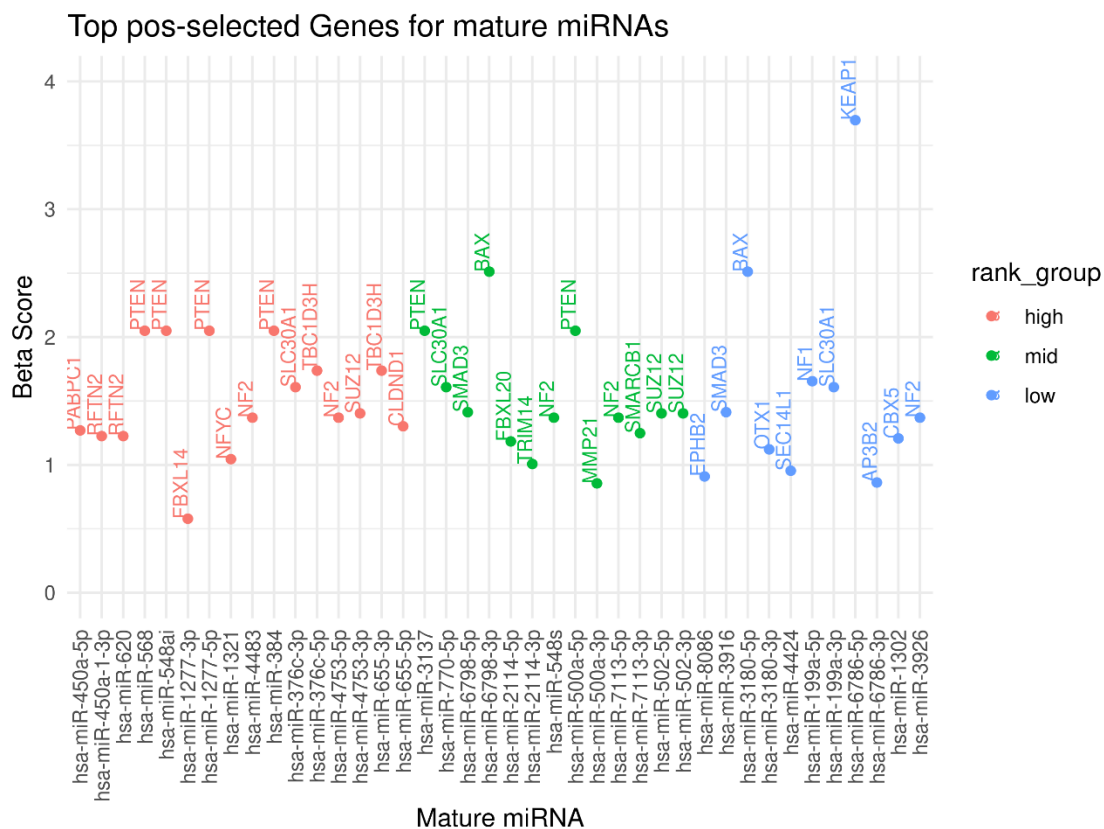


Εικόνα 12: Κατανομές των β scores των θετικά επιλεγμένων γονιδίων-στόχων του κάθε miRNA ανάλογα με τη θέση που διαθέτει στην κατάταξη. Το γράφημα A περιλαμβάνει τις κατανομές των β scores των θετικά επιλεγμένων γονιδίων-στόχων αντιστοιχώντας τους στις πρόδρομες μορφές των miRNAs που έχουν απομονωθεί από κάθε κατάταξη, ενώ το γράφημα B περιλαμβάνει τις κατανομές μόνο των 2% υψηλότερων β scores. Στα γραφήματα C και D γίνεται η μετάβαση από τις πρόδρομες μορφές των miRNAs στις ώριμες και παρουσιάζονται οι αντίστοιχες κατανομές. Οι χρωματισμοί των miRNAs έχουν γίνει με βάση την θέση τους στην αρνητική κατάταξη (negative ranking): τα ροζ είναι στην υψηλότερη θέση (high), τα πράσινα είναι στη μέση (mid) και τα γαλάζια είναι στην κατώτερη θέση (low).

Top pos-selected Genes for mature miRNAs



Εικόνα 13: Διάγραμμα διασποράς των πιο θετικά επιλεγμένων γονιδίων στόχων για τα miRNAs από την ανάλυση RRA του συνόλου δεδομένων Lapatinib Treatment. Το διάγραμμα περιλαμβάνει τον βαθμό θετικής κατάταξης (positive rank) του γονιδίου στόχου με την υψηλότερη θέση στην θετική κατάταξη για το κάθε ώριμο miRNA από τα δείγματα υψηλής (high), μεσαίας (mid) και χαμηλής (low) κατάταξης.



Εικόνα 14: Διάγραμμα διασποράς των πιο θετικά επιλεγμένων γονιδίων στόχων για τα miRNAs από την ανάλυση MLE του συνόλου δεδομένων Laratinib Treatment. Το διάγραμμα περιλαμβάνει το β score του γονιδίου στόχου με το μεγαλύτερο β score για το κάθε ώριμο miRNA από τα δείγματα υψηλής (high), μεσαίας (mid) και χαμηλής (low) κατάταξης.

3.4 Συνδυαστική ανάλυση σημαντικά επιλεγμένων miRNAs και στόχων

Ακολουθως, η ανάλυσή μας επικεντρώθηκε στα στατιστικώς σημαντικά αρνητικά επιλεγμένα miRNAs, τα οποία ρυθμίζουν στατιστικώς σημαντικά θετικά επιλεγμένα γονίδια. Μέσω των χαρτών θερμότητας μπορούμε να απεικονίσουμε και να συγκρίνουμε αυτές τις αλληλεπιδράσεις. Για τα αποτελέσματα RRA και MLE, από τα δεδομένα ATRX Knockout και Laratinib Treatment, έχοντας απομονώσει μόνο τα στατιστικώς σημαντικά αρνητικά επιλεγμένα miRNAs απεικονίσαμε συγκριτικά, μέσω χαρτών θερμότητας, ποια από αυτά ρυθμίζουν στατιστικώς σημαντικά θετικά επιλεγμένα γονίδια και ποια όχι (Εικόνα 15, Εικόνα 16, Εικόνα 17, στις υποενότητες παρακάτω). Στα αποτελέσματα από τα δεδομένα Olaparib Treatment η κατεύθυνση επιλογής που διαλέξαμε για τα γονίδια και miRNAs δεν ευνοούταν από την πειραματική-βιολογική συνθήκη, με αποτέλεσμα να είναι αδύνατον να απομονώσουμε στατιστικώς σημαντικά γονίδια/miRNAs.

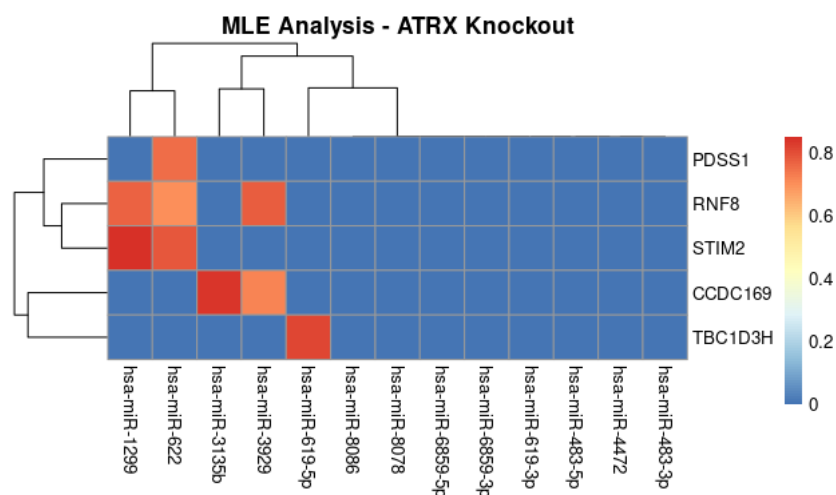
3.4.1 Σύνολο δεδομένων ATRX Knockout

Από την RRA ανάλυση του ATRX Knockout συνόλου δεν προέκυψαν στατιστικά σημαντικά γονίδια με υψηλή βαθμολογία θετικής επιλογής `ros|rank`, οπότε επικεντρωθήκαμε στα ευρήματα της MLE ανάλυσης. Από την MLE ανάλυση του ATRX Knockout συνόλου διατηρήσαμε:

(α) miRNAs (n = 13) απαραίτητα για την αύξηση των ATRX Knockout κυττάρων (β score < 0, FDR < 0.05)

(β) γονίδια (n = 5) των οποίων η παρουσία αναστέλλει την αύξηση των ATRX Knockout κυττάρων (β score > 0, FDR < 0.05)

Από το χάρτη θερμότητας (Εικόνα 15) αναδεικνύονται στατιστικώς σημαντικές οι αλληλεπιδράσεις: α) του miR-1299 με τα γονίδια RNF8 και STIM2, β) του miR-622 με τα PDSS1, RNF8 και STIM2, γ) του miR-3135b με το CCDC169, δ) του miR-3929 με τα RNF8 και CCDC169 και ε) του miR-619-5p με το TBC1D3H.



Εικόνα 15: Χάρτης θερμότητας των πιο στατιστικώς σημαντικών miRNAs και των πιο στατιστικώς σημαντικών γονιδίων που στοχεύουν, από την ανάλυση MLE του ATRX Knockout. Στο διάγραμμα απεικονίζεται η αντιστοίχιση των πιο στατιστικώς σημαντικών αρνητικά επιλεγμένων miRNAs με τους πιο στατιστικώς σημαντικούς θετικά επιλεγμένους γονιδιακούς στόχους τους. Η χρωματική διαβάθμιση αφορά το score πρόβλεψης για την κάθε αλληλεπίδραση. Για τα miRNAs που δεν βρέθηκε στατιστικώς σημαντική αλληλεπίδραση, το score πρόβλεψης ισούται με μηδέν.

Τα στατιστικώς σημαντικά επιλεγμένα γονίδια αποτελούν πιθανοί ογκοκαταστολείς και εμπλέκονται σε βιολογικές διεργασίες που αναστέλλουν την ανάπτυξη και τον πολλαπλασιασμό των κυττάρων ηπατοκυτταρικού καρκινώματος με αποσιώπηση ATRX. Παρουσιάζονται παρακάτω βιβλιογραφικά στοιχεία για τα εν λόγω γονίδια.

- Το γονίδιο TBC1D3H κωδικοποιεί μια πρωτεΐνη, μέλος της TBC1 οικογένειας πρωτεϊνών, η οποία εμπλέκεται στην ενεργοποίηση της δράσης των ενζύμων

GTPase. Αλλαγές στη μεθυλίωση του γονιδίου TBC1D3H και μείωση στην έκφρασή του έχουν συνδεθεί στενά με την εμφάνιση και την ανάπτυξη ηπατοκυτταρικού καρκινώματος^[50].

- Το γονίδιο CCDC169 (Coiled-Coil Domain Containing 169) έχει χαρακτηριστεί ότι σχετίζεται με την εμφάνιση θρόμβωσης και έχει βρεθεί με μειωμένη έκφραση σε δείγματα ιστών γαστρικού αδενοκαρκινώματος^[51].
- Το γονίδιο STIM2 είναι μέλος της οικογένειας μορίων στρωματικών αλληλεπιδράσεων STIM. Διαταραχές στην έκφραση και τη λειτουργία των μορίων STIM έχουν συνδεθεί με την εμφάνιση και τη διατήρηση φαινοτύπων όγκου σε πολλές κακοήθειες, όπως καρκίνοι προστάτη, μαστού, νεφρού, οισοφάγου, δέρματος, εγκεφάλου, παχέος εντέρου, πνεύμονα και ήπατος. Πιο συγκεκριμένα, η αυξημένη έκφραση του STIM2 στο χολαγγειοκαρκίνωμα καταστέλλει την μετάστασή του^[52].
- Το γονίδιο RNF8 κωδικοποιεί την Ring Finger πρωτεΐνη 8, η οποία είναι μια λιγάση E3 ζωτικής σημασίας για την επιδιόρθωση του DNA. Μελέτες έχουν δείξει ότι το RNF8 εμπλέκεται επίσης σε πολλές βιολογικές διεργασίες που σχετίζονται με τον καρκίνο, όπως η ογκογένεση και η μετάσταση^[53].
- Το γονίδιο PDSS1 κωδικοποιεί το ένζυμο Decaprenyl-diphosphate synthase subunit 1, το οποίο σχετίζεται με τη σύνθεση του συνενζύμου Q. Έχει χαρακτηριστεί ως προγνωστικός δείκτης για το ηπατοκυτταρικό καρκίνωμα και εμπλέκεται σε πολλά μονοπάτια σχετικά με τον κυτταρικό κύκλο^[54].

Βάσει του τύπου επιλογής που έχουν τα στατιστικώς σημαντικά επιλεγμένα miRNAs, επιδρούν προς όφελος των κυττάρων ηπατοκυτταρικού καρκινώματος με αποσιώπηση ATRX, ρυθμίζοντας αρνητικά την έκφραση γονιδίων που αναστέλλουν την κυτταρική επιβίωση και/ή αύξηση. Τα συγκεκριμένα miRNAs αποτελούν πιθανά ζεύγη συνθετικής θνησιμότητας με το γονίδιο ATRX, καθώς η ταυτόχρονη αποσιώπησή τους προάγει την κυτταρική απόπτωση. Πραγματοποιώντας βιβλιογραφική αναζήτηση, βρήκαμε έρευνες που σχολιάζουν τη λειτουργία των παραπάνω miRNAs στο ηπατοκυτταρικό καρκίνωμα και/ή σε κάποιον άλλο τύπο καρκίνου.

- Το miR-1299 αναφέρεται εκτενώς ως ογκοκαταστολέας με χαμηλή έκφραση στον ηπατοκυτταρικό καρκίνο. Ρυθμίζει την έκφραση γονιδίων που σχετίζονται με βιολογικές διεργασίες όπως ο κυτταρικός πολλαπλασιασμός, η μετανάστευση, η επιβίωση και ο προγραμματισμένος κυτταρικός θάνατος^[55]. Ενδεχομένως η επιλογή που εμφανίζει στα πειραματικά μας δεδομένα να υποδηλώνει ότι στη συνθήκη ελέγχου έχει πιο έντονη ογκοκατασταλτική δράση, ενώ κατά την αποσιώπηση ATRX αυτή μετριάζεται.
- Τα επίπεδα έκφρασης του miR-622 διαταράσσονται σε πολλούς τύπους καρκίνου. Λειτουργικά, μπορεί να αναστείλει τον κυτταρικό πολλαπλασιασμό και να προάγει την κυτταρική απόπτωση^[56]. Ωστόσο, μελέτες υποδηλώνουν ότι

μπορεί να έχει και χαρακτήρα ευνοϊκό για την ανάπτυξη και ανθεκτικότητα των καρκινικών κυττάρων^[57].

- Το miR-3135b έχει βρεθεί με χαμηλά επίπεδα έκφρασης σε άλλους καρκίνους, όπως αυτός του παχέος εντέρου, και η υπερέκφρασή του αναστέλλει τον κυτταρικό πολλαπλασιασμό και ενεργοποιεί την κυτταρική απόπτωση^[58]. Σε άλλες περιπτώσεις, όπως στον τριπλά αρνητικό καρκίνο του μαστού αναφέρεται ότι μπορεί να φέρει ογκογενετικές λειτουργίες, επάγοντας μεταστατικές διεργασίες (μετασχηματισμός κυττάρων από επιθηλιακό χαρακτήρα σε μεσεγχυματικό, EMT)^[59].
- Το miR-3929 αναστέλλει τον πολλαπλασιασμό και επάγει την απόπτωση των καρκινικών κυττάρων του τραχήλου της μήτρας^[60].
- Το miR-619-5p έχει δείξει ογκοκατασταλτική δράση στο πλακώδες καρκίνωμα κεφαλής και λαιμού (squamous head and neck cancer)^[61] και έχει χαρακτηριστεί ο κατασταλτικός του ρόλος στον επιθετικό πολλαπλασιασμό των κυττάρων καρκινώματος παχέος εντέρου^[62]. Αντιθέτως, αναφέρονται επαγωγικοί του ρόλοι στη μετάσταση και την καρκινική αγγειογένεση στο μη-μικροκυτταρικό καρκίνο του πνεύμονα^[63].

3.4.2 Σύνολο δεδομένων επώασης με Lapatinib

Από τις αναλύσεις RRA και MLE του συνόλου που επώαστηκε με Lapatinib διατηρήθηκαν:

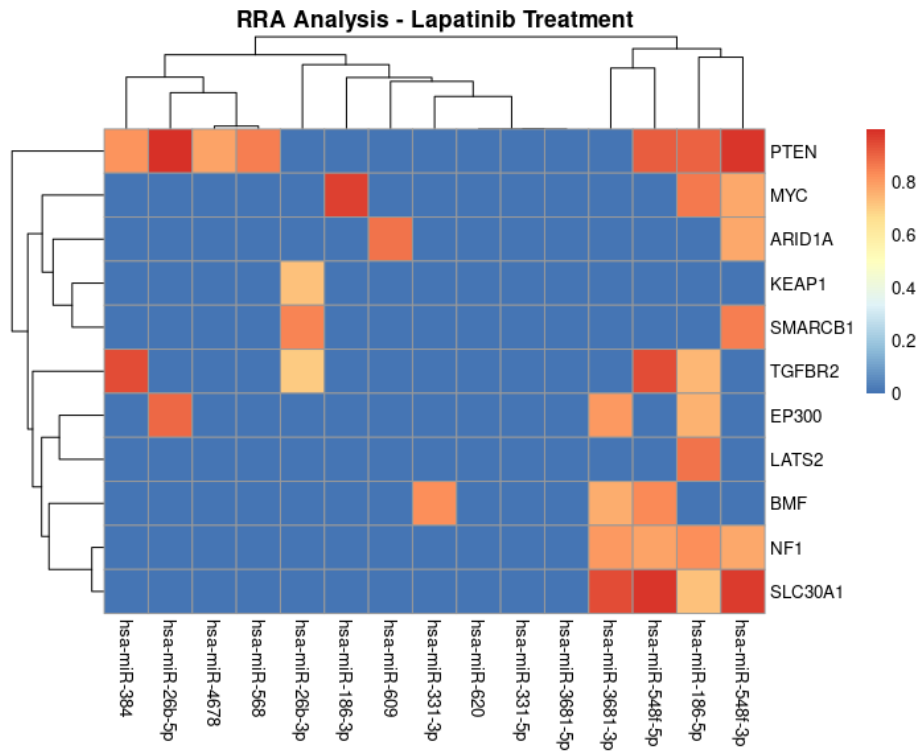
(α) miRNAs απαραίτητα για την αύξηση των κυττάρων με Lapatinib ($n_{RRA} = 15$, $FDR < 0.47$, $n_{MLE} = 16$, β score < 0 , $FDR < 0.25$)

(β) γονίδια των οποίων η απουσία προσδίδει ανθεκτικότητα στα κύτταρα με Lapatinib ($n_{RRA} = 11$, $FDR < 0.05$, $n_{MLE} = 2$, β score > 0 , $FDR < 0.36$)

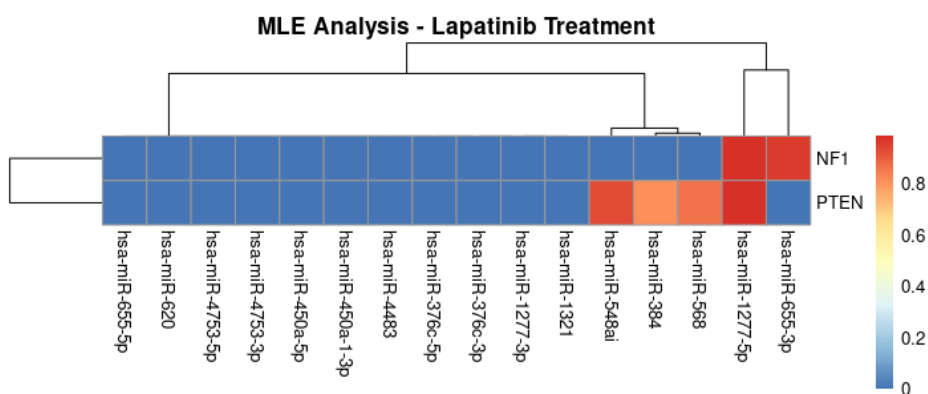
Και από τις δυο αναλύσεις (RRA και MLE), αναδείχθηκαν απαραίτητα miRNAs που στοχεύουν και ρυθμίζουν επιβλαβή (deleterious) γονίδια. Ενδεικτικά, κάποιες από αυτές τις αλληλεπιδράσεις είναι:

RRA ανάλυση: α) η στόχευση των miR-384, -26b-5p, -4678, -568, -548f-5p/3p και -186-5p έναντι του μεταγράφου PTEN, β) των miR-186-5p/3p και -548f-3p έναντι του μεταγράφου MYC, γ) των miR-3681-3p, -548f-5p/3p και -186-5p έναντι των NF1 και SLC30A1, δ) των miR-384, -25b-3p, -548f-5p και -186-5p έναντι του TGFBR2 (αναλυτικότερα στο χάρτη θερμότητας στην **Εικόνα 16**)

MLE ανάλυση: α) η στόχευση των miR-548ai, -384, -568 και -1277-5p στο γονίδιο PTEN και β) των miR-655-3p και -1277-5p στο γονίδιο NF1 (**Εικόνα 17**)



Εικόνα 16: Χάρτης θερμότητας των πιο σημαντικών miRNAs και γονιδίων-στόχων τους από την ανάλυση RRA του Lapatinib Treatment. Στο διάγραμμα απεικονίζεται η αντιστοίχιση των πιο στατιστικώς σημαντικών αρνητικά επιλεγμένων miRNAs με τους πιο στατιστικώς σημαντικούς θετικά επιλεγμένους γονιδιακούς στόχους τους. Η χρωματική διαβάθμιση αφορά το score πρόβλεψης για την κάθε αλληλεπίδραση. Για τα miRNAs που δεν βρέθηκε στατιστικώς σημαντική αλληλεπίδραση, το score πρόβλεψης ισούται με μηδέν.



Εικόνα 17: Χάρτης θερμότητας των πιο σημαντικών miRNAs και γονιδίων-στόχων τους από την ανάλυση MLE του Lapatinib Treatment. Στο διάγραμμα απεικονίζεται η αντιστοίχιση των πιο στατιστικώς σημαντικών αρνητικά επιλεγμένων miRNAs με τους πιο στατιστικώς σημαντικούς θετικά επιλεγμένους γονιδιακούς στόχους τους. Η χρωματική διαβάθμιση αφορά το score πρόβλεψης για την κάθε αλληλεπίδραση. Για τα miRNAs που δεν βρέθηκε στατιστικώς σημαντική αλληλεπίδραση, το score πρόβλεψης ισούται με μηδέν.

Τα στατιστικώς σημαντικά επιλεγμένα γονίδια παρουσιάζονται από τη δοκιμασία ως επιβλαβή-θνησιγόνα κατά την επώαση με το θεραπευτικό παράγοντα Laratinib. Με άλλα λόγια, σε μια θεραπεία με Laratinib αναμένεται καρκινικά κύτταρα που τα υπερ-εκφράζουν να αποκρίνονται σημαντικά στη θεραπεία και να μην επιβιώνουν. Παρουσιάζονται παρακάτω γνωστές πληροφορίες για τα γονίδια αυτά και τη συσχέτισή τους με τον καρκίνο.

- Το PTEN (Phosphatase and Tensin Homolog) είναι ένα ογκοκατασταλτικό γονίδιο, του οποίου οι μεταλλάξεις έχουν συσχετιστεί με μεγάλο αριθμό καρκινικών τύπων. Μελέτες έχουν δείξει ότι η απενεργοποίησή του συνδέεται στενά με την εξέλιξη και την απόκριση σε χημειοθεραπείες του γαστρικού καρκίνου^[64].
- Το γονίδιο NF1 (Neurofibromin 1) αποτελεί ογκοκατασταλτικό παράγοντα και η απώλεια της έκφρασής του έχει συνδεθεί με την ανάπτυξη και την εξέλιξη καρκίνων. Πιο συγκεκριμένα, μεταλλάξεις στο NF1, αλλαγή του αριθμού των αντιγράφων του ή ρύθμιση μέσω miRNAs μπορούν να οδηγήσουν σε διάφορους φαινοτύπους καρκίνου, όπως η μειωμένη απόπτωση και ο αυξημένος πολλαπλασιασμός^[65].
- Το γονίδιο SLC30A1 (Solute Carrier Family 30 Member 1), έχει βρεθεί ότι υπερ-εκφράζεται σημαντικά σε ιστούς γαστρικού καρκίνου και σχετίζεται με το στάδιο του όγκου^[66].
- Το MYC (C-MYC) είναι ένα εντατικά μελετημένο ογκογονίδιο και σχετίζεται με αρκετούς τύπους καρκίνου. Η απορρύθμιση του MYC έχει συσχετιστεί κυρίως με κακή/φτωχή πρόγνωση της πορείας της νόσου και την ενίσχυση μονοπατιών που εμπλέκονται στην γαστρική καρκινογένεση^[67].
- Η μείωση στην έκφραση του γονιδίου TGFBR2 (Transforming Growth Factor Beta Receptor 2) έχει συσχετιστεί στενά με το μέγεθος του όγκου, την πιθανότητα μετάστασης και την προγνωστική δυνατότητα του καρκίνου του μαστού^[68].

Αντίστοιχα, από τη δοκιμασία αναδείχθηκαν σημαντικά επιλεγμένα miRNAs, τα οποία επιδρούν ευεργετικά για την επιβίωση κυττάρων κατά την επώαση με Laratinib, δηλαδή η υπερέκφρασή τους διαφαίνεται πως θα προσδίδει στα καρκινικά κύτταρα ανθεκτικότητα στη δράση του Laratinib. Παρακάτω παρατίθενται βιβλιογραφικά στοιχεία για κάποια από τα εν λόγω miRNAs.

- Το miR-548f-3p έχει εντοπιστεί με διαφοροποιημένα επίπεδα έκφρασης και στο πλακώδες καρκίνωμα κεφαλής και λαιμού και αναδείχθηκε ως δυνητικός προγνωστικός δείκτης για την πρόβλεψη της επιβίωσης (Overall Survival Prediction)^[69].
- Για το miR-186-5p έχει αναφερθεί ότι εκφράζεται έκτοπα σε διάφορους καρκίνους και βρέθηκε ότι εμπλέκεται σε διαφορετικές λειτουργίες των καρκινικών κυττάρων, όπως ο πολλαπλασιασμός και η μετάσταση. Στον καρκίνο του παχέος εντέρου αποδείχθηκε ότι η υπερέκφρασή του αναστέλλει τον

κυτταρικό πολλαπλασιασμό και την μετάσταση (στοχεύοντας στο γονίδιο ZEB1)^[70].

- Η απορρύθμιση της έκφρασης του miR-3681-3p παρατηρήθηκε σε καρκινικά κύτταρα τραχήλου της μήτρας, όπου αυτό συσχετίστηκε με την αρνητική ρύθμιση του πολλαπλασιασμού των κυττάρων και με την κυτταρική απόπτωση^[71].
- Οι λειτουργίες του miR-384 έχουν μελετηθεί σε διάφορους τύπους καρκίνων, συμπεριλαμβανομένου του θηλώδους καρκίνου του θυρεοειδούς^[72], του παγκρεατικού καρκίνου^[73], του μη-μικροκυτταρικού καρκίνου του πνεύμονα^[74] και του καρκίνου του παχέος εντέρου^[75]. Σε αυτές τις περιπτώσεις, το miR-384 βρέθηκε να στοχεύει γονίδια ευνοϊκά για την ανάπτυξη και μετάσταση του όγκου.
- Σε προηγούμενες μελέτες, η δράση του miR-1277-5p έχει ερευνηθεί με μειωμένα επίπεδα έκφρασης στον γαστρικό καρκίνο^[76] και στον καρκίνο του παχέος εντέρου^[77].

3.5 Κοινά εμπλουτισμένα μονοπάτια KEGG από επιλεγμένα miRNAs και γονίδια

Συλλέχθηκαν αποτελέσματα από τις αναλύσεις εμπλουτισμού μονοπατιών για γονίδια και miRNAs. Συγκεκριμένα, από τις αναλύσεις:

(α) αλγόριθμος RRA: γονιδίων με θετική βαθμολογία ($p < 0.05$) – ανάλυση μονοπατιών MAGeCKFlute

(β) αλγόριθμος MLE: γονιδίων με θετικό β score, έπειτα από φιλτράρισμα (**Εικόνα 4**) – ανάλυση μονοπατιών MAGeCKFlute

(γ) αλγόριθμος RRA: σημαντικά επιλεγμένων miRNAs με αρνητική βαθμολογία – ανάλυση μονοπατιών miRPath

(δ) αλγόριθμος MLE: σημαντικά επιλεγμένων miRNAs με αρνητικό β score – ανάλυση μονοπατιών miRPath

Σε κάθε ζεύγος αποτελεσμάτων (εμπλουτισμένα μονοπάτια από miRNAs και γονίδια που προέκυψαν από την ίδια μεθοδολογία: RRA ή MLE), εντοπίστηκαν και απεικονίστηκαν μόνο κοινά μονοπάτια, που ενισχύονταν τόσο από επιλεγμένα γονίδια όσο και από miRNAs. Από αυτή τη διαδικασία πήραμε αποτελέσματα τόσο για τα RRA όσο και για τα MLE δεδομένα από τα δύο σύνολα δεδομένων (ATRX Knockout και Lapatinib Treatment), τα απεικονίσαμε και τα συγκρίναμε μεταξύ τους (**Εικόνα 18**, **Εικόνα 19**). Στις επόμενες ενότητες παρουσιάζονται και σχολιάζονται ενδεικτικά επιλεγμένα αποτελέσματα, με γνώμονα τη βιολογική σημασία των μονοπατιών και τη στατιστική σημαντικότητα.

3.5.1 Εμπλουτισμένα μονοπάτια κατά την ATRX αποσιώπηση

Σχετικά με το σύνολο ATRX Knockout επιλέξαμε τα μονοπάτια KEGG “PD-L1 expression and PD-1 checkpoint pathway in cancer” και “Neurotrophin signaling pathway”, από τα κοινά που προέκυψαν από τις αναλύσεις RRA, ενώ από τα κοινά που προέκυψαν από τις αναλύσεις MLE, επιλέξαμε το μονοπάτι “Prolactin signaling pathway”.

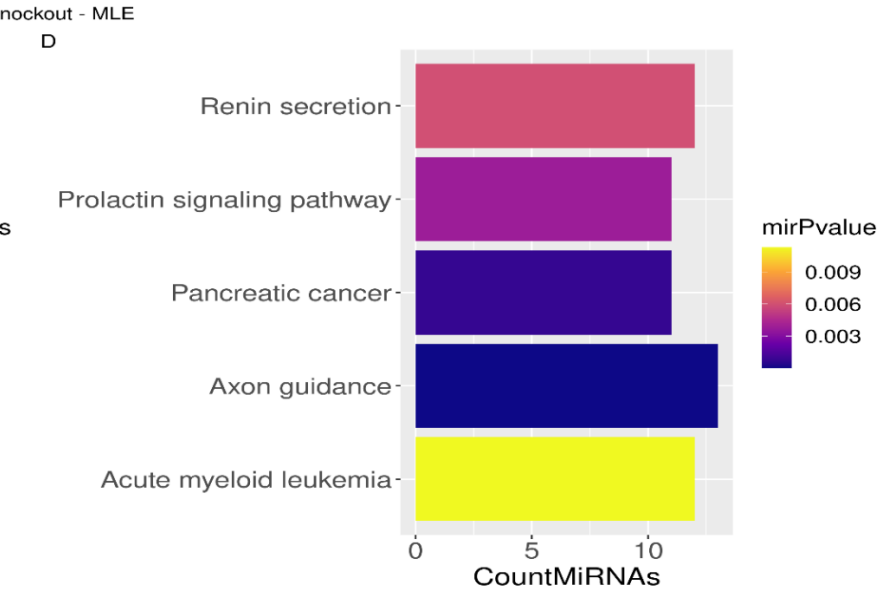
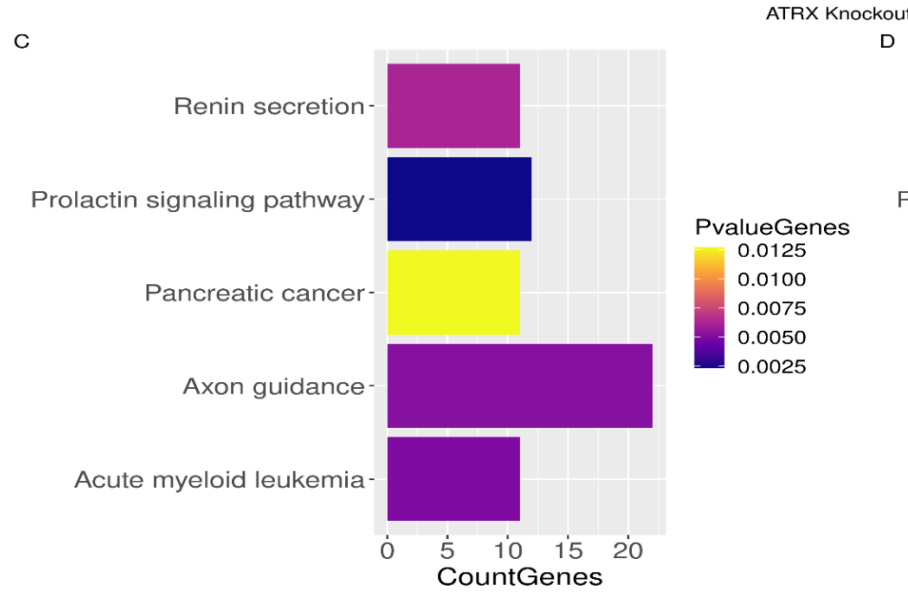
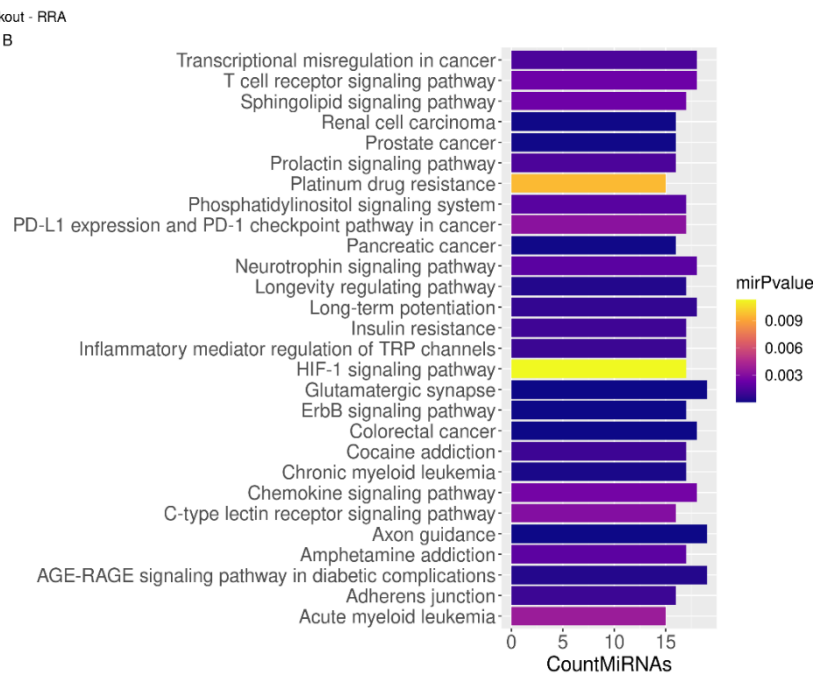
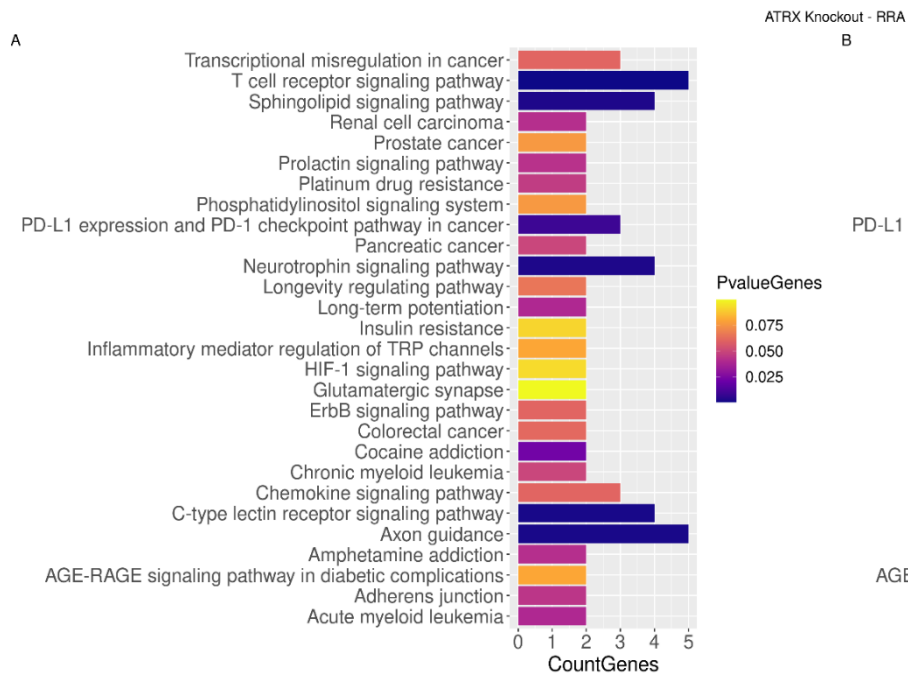
- **PD-L1 expression and PD-1 checkpoint pathway in cancer:** Η πρωτεΐνη προγραμματισμένου κυτταρικού θανάτου 1 (PD-1) και το μόριο-προσδέτης της (PD-L1) σε φυσιολογικά κύτταρα αποτελούν ρυθμιστικά μόρια του ανοσοποιητικού που διατηρούν την αυτό-ανοχή στον οργανισμό ρυθμίζοντας τον βαθμό ενεργοποίησης των T και B λεμφοκυττάρων μεταξύ άλλων τύπων ανοσοκυττάρων. Το μονοπάτι PD-1/PD-L1 αξιοποιείται από καρκινικά κύτταρα τα οποία αυξάνουν την έκφραση της PD-L1 στην επιφάνειά τους, ώστε αυτά να διαφεύγουν της αναγνώρισης από το ανοσοποιητικό σύστημα. Ο προσδέτης (ligand) PD-L1 συνδέεται στον υποδοχέα PD-1 των T λεμφοκυττάρων και ενεργοποιεί τη μεταγωγή σήματος PD-L1/PD-1, αναστέλλοντας την αντικαρκινική δράση των T λεμφοκυττάρων. Ερευνώντας τα παραπάνω, έχουν αναπτυχθεί πολλά υποσχόμενες καρκινικές ανοσοθεραπείες, που στηρίζονται στην παρεμπόδιση (blockade) του PD-1/PD-L1 μονοπατιού^[78, 79]. Στην απεικόνιση του μονοπατιού (**Εικόνα 20**), χρωματίζονται τα γονίδια που συμμετέχουν στο μονοπάτι και είναι σημαντικά στην θετική επιλογή (τα κύτταρα που είχαν αποσιώπηση στα συγκεκριμένα γονίδια πολλαπλασιάστηκαν γρηγορότερα).
- **Neurotrophin signaling pathway:** Οι νευροτροφίνες αποτελούν μια οικογένεια πρωτεϊνών, οι οποίες στα νευρικά κύτταρα ρυθμίζουν διάφορες λειτουργίες που σχετίζονται με την επιβίωση, την ανάπτυξη και τη λειτουργία. Στις πρωτεΐνες αυτές ανήκουν ο νευροαυξητικός παράγοντας (NGF), ο προερχόμενος από τον εγκέφαλο νευροτροφικός παράγοντας (BDNF), η νευροτροφίνες 3 (NT-3) και 4 (NT-4). Οι νευροτροφίνες και οι υποδοχείς νευροτροφίνης βρίσκονται και σε καρκινικά και κύτταρα του στρώματος (stromal cells) και συνδέονται με πολλά είδη όγκου. Ο NGF και ο BDNF συνδέονται με την ανάπτυξη όγκου, προάγοντας τον πολλαπλασιασμό, την αγγειογένεση και την εισβολή^[80]. Επίσης, για την νευροτροφίνη 3 έχει βρεθεί ότι αποτελεί ογκοκατασταλτική πρωτεΐνη στο ηπατοκυτταρικό καρκίνωμα^[81]. Για το συγκεκριμένο μονοπάτι έχει αναδειχθεί ότι γονίδια που συμμετέχουν σε αυτό είναι υπερμεθυλιωμένα και απορρυθμίζονται στο ηπατοκυτταρικό καρκίνωμα^[80]. Πιθανόν τα miRNAs που έχουν βρεθεί σημαντικά για τον πολλαπλασιασμό των κυττάρων ηπατοκυτταρικού καρκινώματος με αποσιώπηση ATRX να ρυθμίζουν αρνητικά κάποια από αυτά τα γονίδια (με γαλάζιο περίγραμμα στην **Εικόνα 21**).
- **Prolactin signaling pathway:** Η προλακτίνη είναι μια πολυπεπτιδική ορμόνη που εμπλέκεται σε ένα ευρύ φάσμα βιολογικών λειτουργιών (ωσμωρύθμιση, γαλακτοφορία, αναπαραγωγή, ανάπτυξη, μεταβολισμός και ρύθμιση ανοσοαπόκρισης). Πειράματα (σε ποντίκια) έχουν δείξει ότι η προλακτίνη μπορεί να έχει ανασταλτική δράση έναντι του ηπατοκυτταρικού καρκινώματος^[82]. Στην

απεικόνιση του μονοπατιού KEGG (**Εικόνα 22**) φαίνεται η ρύθμιση που ασκούν miRNAs (στατιστικώς σημαντικά στην αρνητική επιλογή – ευεργετικά για τον πολλαπλασιασμό των κυττάρων ηπατοκυτταρικού καρκινώματος με αποσιώπηση ATRX) σε γονίδια που συμμετέχουν σε αυτό. Η αποσιώπηση αυτών των miRNAs έχει θνησιγόνο επίδραση στα κύτταρα ηπατοκυτταρικού καρκινώματος με αποσιώπηση ATRX, πιθανόν επειδή ρυθμίζουν αρνητικά την έκφραση γονιδίων που συμμετέχουν στη σηματοδότηση της προλακτίνης.

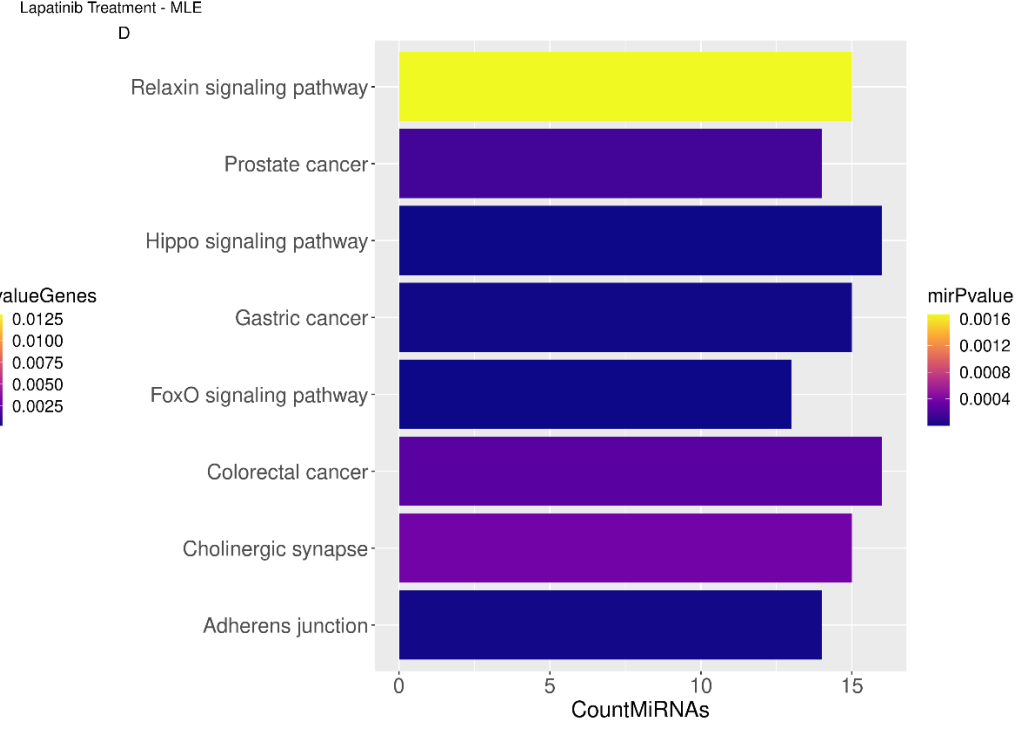
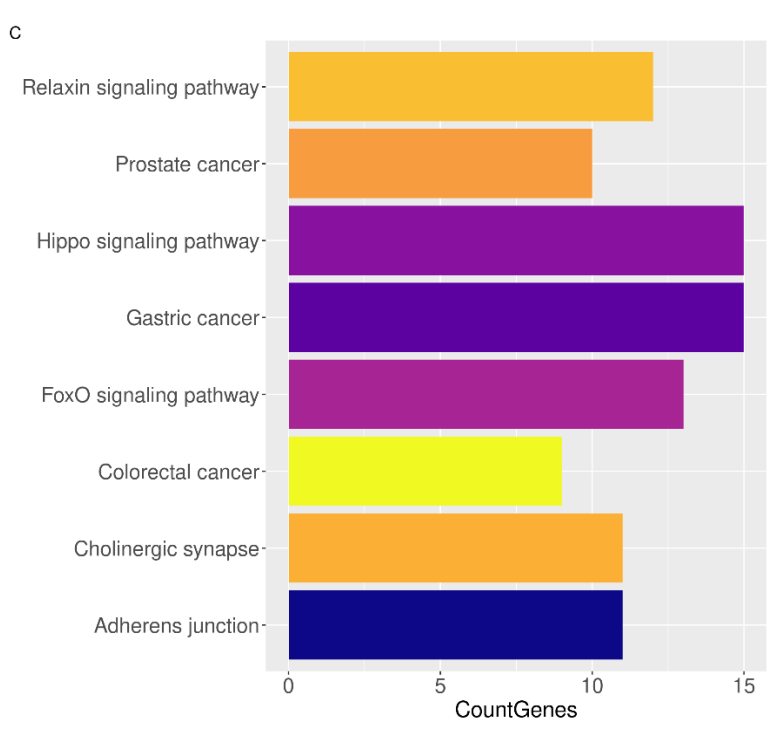
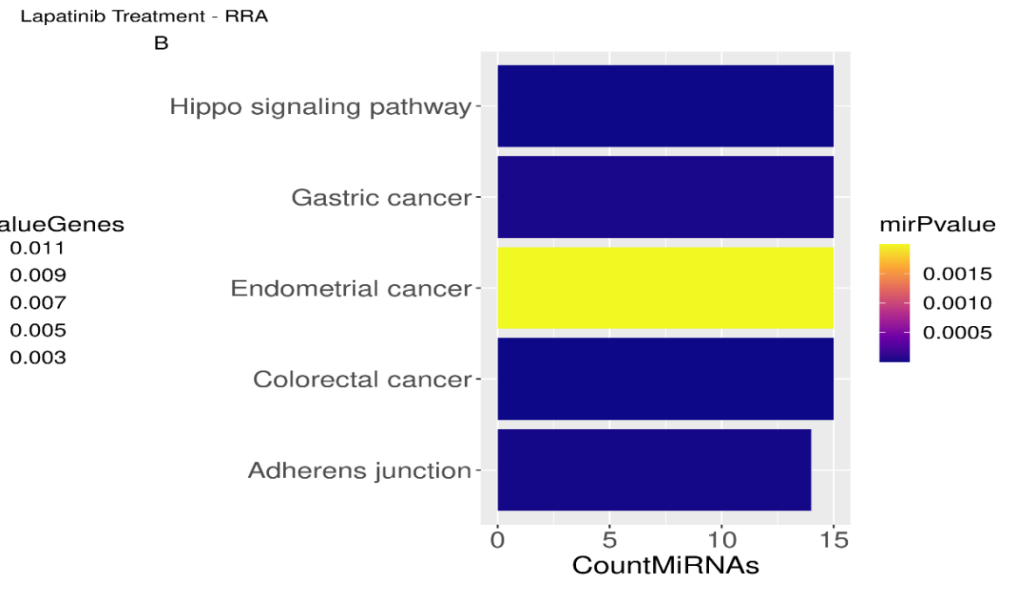
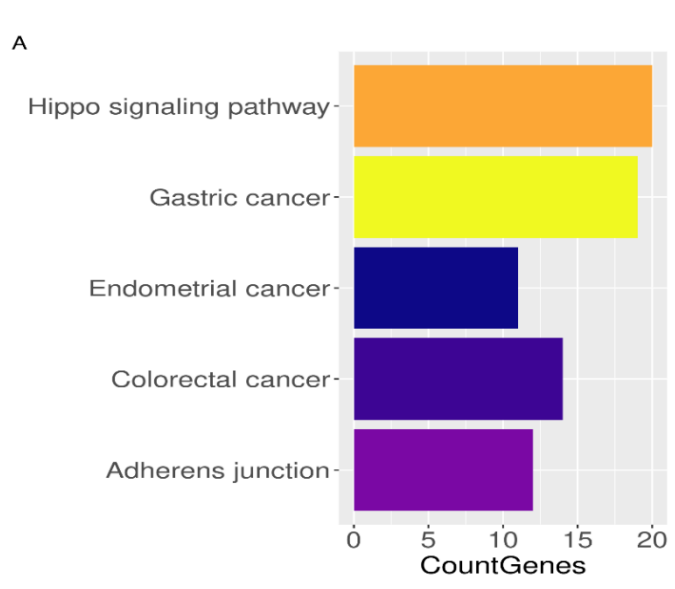
3.5.2 Εμπλουτισμένα μονοπάτια κατά επώαση με *Laratinib*

Για το σύνολο *Laratinib* Treatment επιλέξαμε το μονοπάτι KEGG “Adherens junction” από τα κοινά των αναλύσεων RRA και το “Gastric cancer” από αυτά των αναλύσεων MLE.

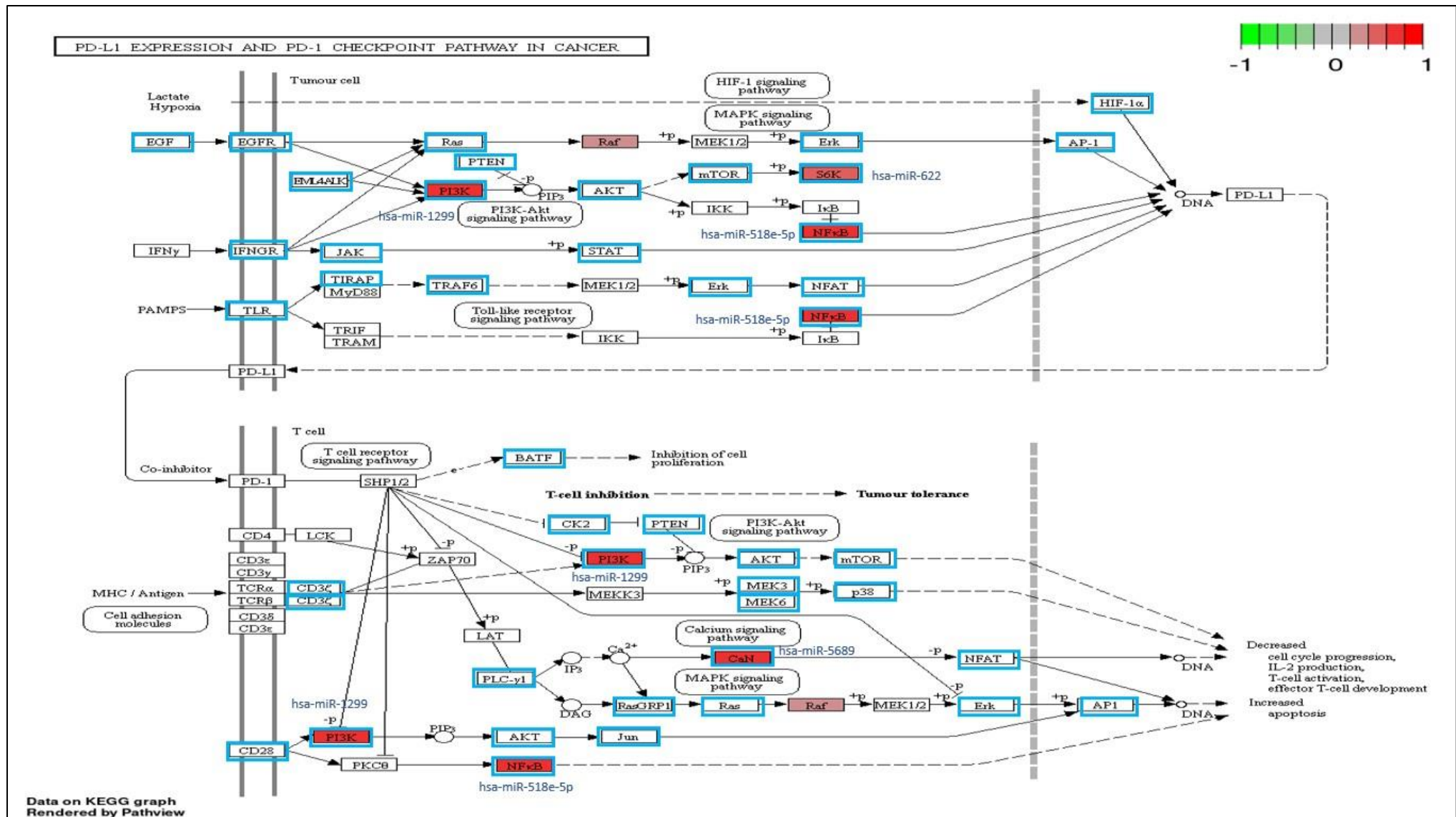
- **Adherens junction:** Οι συνδέσεις προσκόλλησης κυττάρου-κυττάρου (AJs) αποτελούν τον πιο κοινό τύπο διακυτταρικής προσκόλλησης, είναι σημαντικές για τη διατήρηση της αρχιτεκτονικής των ιστών και της πολικότητας των κυττάρων και μπορούν να περιορίσουν την κίνηση και τον πολλαπλασιασμό των κυττάρων. Πολλά μόρια που συμμετέχουν στις AJs καταλαμβάνουν ενεργούς ρόλους σε καρκίνους καθώς ρυθμίζουν τον κυτταρικό πολλαπλασιασμό και τη γονιδιακή έκφραση. Διαταραχές στην έκφραση και λειτουργία των Cadherins και των Catenins, σχετίζονται με πολλούς τύπους καρκίνου. Πιο συγκεκριμένα, η E-cadherin φέρει ογκοκατασταλτικό ρόλο στον γαστρικό καρκίνο και η απορρύθμισή της βοηθάει στην ανάπτυξη του όγκου^[83]. Η σίγαση των γονιδίων που έχουν αναδειχθεί σημαντικά στην θετική επιλογή και συμμετέχουν σε αυτό το μονοπάτι ευνοεί την επιβίωση των καρκινικών κυττάρων κατά την επώαση *Laratinib* (ένα από αυτά είναι η E-cadherin). Αυτά τα γονίδια είναι χρωματισμένα στην απεικόνιση του μονοπατιού (**Εικόνα 23**) και ρυθμίζονται από τα επιλεγμένα miRNAs.
- **Gastric cancer:** Ο γαστρικός καρκίνος χωρίζεται σε δύο διακριτές ιστολογικές ομάδες, τον εντερικό και τον διάχυτο τύπο. Στη συγκεκριμένη μελέτη χρησιμοποιήσαμε κύτταρα γάστρο-οισοφαγικού αδενοκαρκινώματος. Στην απεικόνιση του μονοπατιού (**Εικόνα 24**) παρατηρούνται διαταραχές οι οποίες προκαλούν αδενοκαρκίνωμα (εντερικού και διάχυτου τύπου) με χρωματισμένα τα γονίδια (σημαντικά στη θετική επιλογή), των οποίων η σίγαση ευνόησε την επιβίωση των καρκινικών κυττάρων. Κάποια από αυτά, όπως το APC και το CDH1, αποτελούν ογκοκατασταλτικά γονίδια και ρυθμίζονται αρνητικά από miRNAs τα οποία είναι στατιστικώς σημαντικά στην αρνητική επιλογή. Η καταστολή αυτών των miRNAs θα μπορούσε να αυξήσει την έκφραση των ογκοκατασταλτικών γονιδίων και να βελτιώσει την αποτελεσματικότητα της θεραπείας με *Laratinib*.



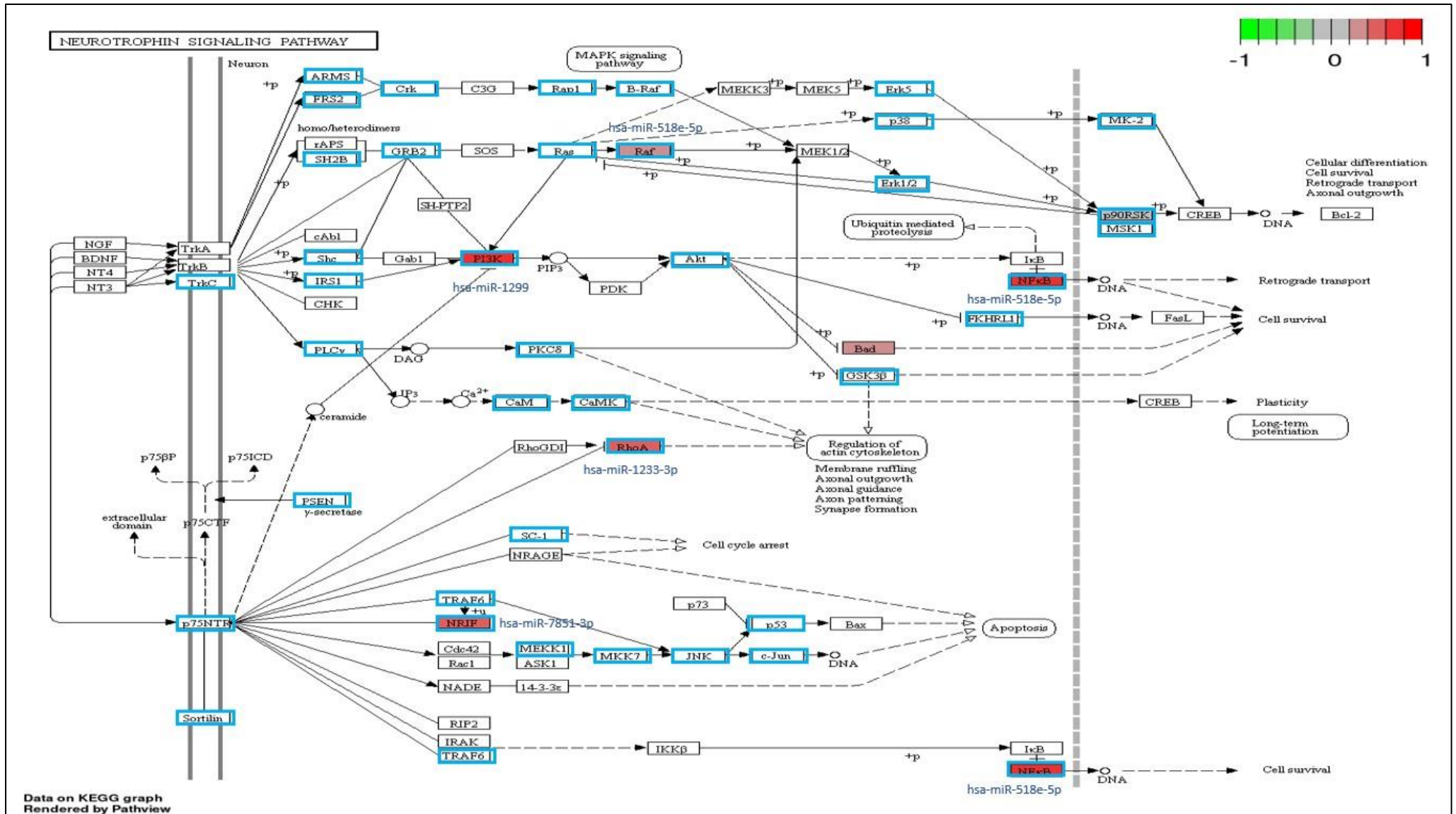
Εικόνα 18: Συγκρίσεις των κοινών μονοπατιών από τις αναλύσεις εμπλουτισμού των επιλεγμένων γονιδίων και miRNAs για το σύνολο δεδομένων ATRX Knockout. Με τα διαγράμματα A και B παρατίθενται, αντίστοιχα, τα χαρακτηριστικά των κοινών μονοπατιών από την γονιδιακή ανάλυση εμπλουτισμού (MAGeCK Flute) και από την ανάλυση εμπλουτισμού των miRNAs (miRPath) που προέκυψαν από τα αποτελέσματα του αλγορίθμου RRA. Με τα διαγράμματα C και D παρατίθενται, με αντίστοιχο τρόπο, αυτά που προέκυψαν από τα αποτελέσματα του αλγορίθμου MLE. Η χρωματική διαβάθμιση συμβολίζει τα p-values που έχουν αποδοθεί στο κάθε μονοπάτι. Στον άξονα x φαίνεται το πλήθος των επιλεγμένων miRNAs/γονιδίων ανά μονοπάτι.



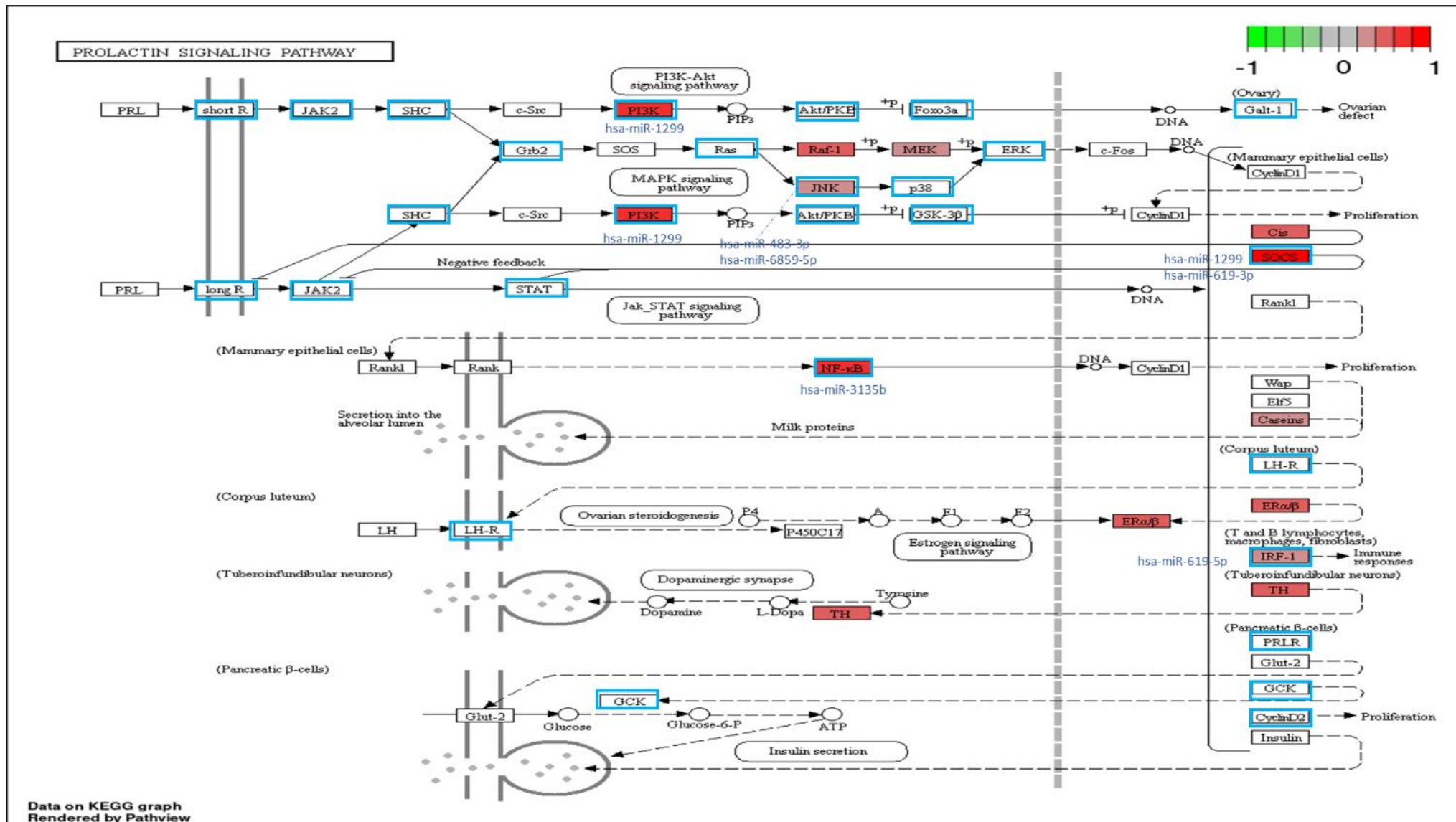
Εικόνα 19: Συγκρίσεις των κοινών μονοπατιών από τις αναλύσεις εμπλουτισμού των επιλεγμένων γονιδίων και miRNAs για το σύνολο δεδομένων Lapatinib Treatment. Με τα διαγράμματα A και B παρατίθενται, αντίστοιχα, τα χαρακτηριστικά των κοινών μονοπατιών από την γονιδιακή ανάλυση εμπλουτισμού (MAGeCK Flute) και από την ανάλυση εμπλουτισμού των miRNAs (miRPath) που προέκυψαν από τα αποτελέσματα του αλγορίθμου RRA. Με τα διαγράμματα C και D παρατίθενται, με αντίστοιχο τρόπο, αυτά που προέκυψαν από τα αποτελέσματα του αλγορίθμου MLE. Η χρωματική διαβάθμιση συμβολίζει τα p-values που έχουν αποδοθεί στο κάθε μονοπάτι. Στον άξονα x φαίνεται το πλήθος των επιλεγμένων miRNAs/γονιδίων ανά μονοπάτι.



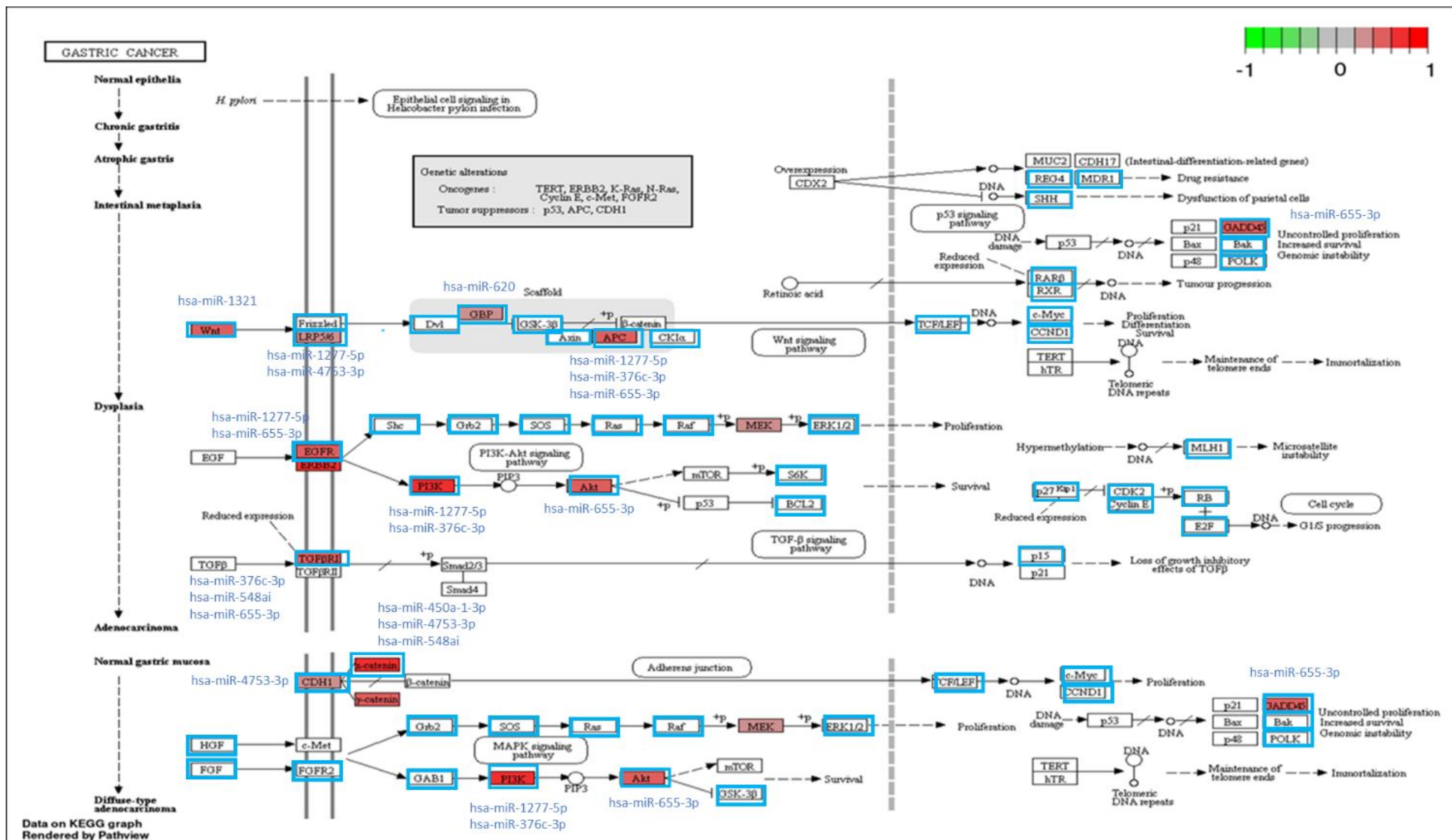
Εικόνα 20: Μονοπάτι έκφρασης του PD-L1 και σηματοδότηση PD-1. Τα γονίδια υπό επιλογή (ανάλυση εμπλουτισμού σε αποτελέσματα RRA) που αναστέλλουν την αύξηση σε ATRX Knockout κύτταρα είναι χρωματισμένα με αποχρώσεις κόκκινου. Η χρωματική διαβάθμιση συμβολίζει την βαθμολογία LFC (Log Fold Change) που έχει αποδοθεί στο κάθε επιλεγμένο γονίδιο. Με γαλάζιο περίγραμμα σημαίνουν τα γονίδια-στόχοι των υπό επιλογή miRNAs (miRNAs που επάγουν αύξηση σε ATRX Knockout). Για τα γονίδια υπό επιλογή είναι επισημειωμένα και τα miRNAs από τα οποία ρυθμίζονται.



Εικόνα 21: Μονοπάτι σηματοδότησης νευροτροφίνης. Τα γονίδια υπό επιλογή (ανάλυση εμπλουτισμού σε αποτελέσματα RRA) που αναστέλλουν την αύξηση σε ATRX Knockout κύτταρα είναι χρωματισμένα με αποχρώσεις κόκκινου. Η χρωματική διαβάθμιση συμβολίζει την βαθμολογία LFC (Log Fold Change) που έχει αποδοθεί στο κάθε επιλεγμένο γονίδιο. Με γαλάζιο περίγραμμα σημαίνονται τα γονίδια-στόχοι των υπό επιλογή miRNAs (miRNAs που επάγουν αύξηση σε ATRX Knockout). Για τα γονίδια υπό επιλογή είναι επισημειωμένα και τα miRNAs από τα οποία ρυθμίζονται.



Εικόνα 22: Μονοπάτι σηματοδότησης προλακτίνης. Τα γονίδια υπό επιλογή (ανάλυση εμπλουτισμού σε αποτελέσματα MLE) που αναστέλλουν την αύξηση σε ATRX Knockout κύτταρα είναι χρωματισμένα με αποχρώσεις κόκκινου. Η χρωματική διαβάθμιση συμβολίζει το β score που έχει αποδοθεί στο κάθε επιλεγμένο γονίδιο. Με γαλάζιο περίγραμμα σημαίνουν τα γονίδια-στόχοι των υπό επιλογή miRNAs (miRNAs που επάγουν αύξηση σε ATRX Knockout κύτταρα). Για τα γονίδια υπό επιλογή είναι επισημειωμένα και τα miRNAs από τα οποία ρυθμίζονται.



Εικόνα 24: Μονοπάτι γαστρικού καρκίνου. Τα γονίδια υπό επιλογή (ανάλυση εμπλουτισμού σε αποτελέσματα MLE) που επάγουν ευαισθησία στο Laratiniib είναι χρωματισμένα με αποχρώσεις κόκκινου. Η χρωματική διαβάθμιση συμβολίζει το β score που έχει αποδοθεί στο κάθε επιλεγμένο γονίδιο. Με γαλάζιο περίγραμμα σημαίνονται τα γονίδια-στόχοι των υπό επιλογή miRNAs (miRNAs που επάγουν ανθεκτικότητα στο Laratiniib). Για τα γονίδια υπό επιλογή είναι επισημειωμένα και τα miRNAs από τα οποία ρυθμίζονται.

4 Συμπεράσματα

Στο πλαίσιο αυτής της πτυχιακής εργασίας, αναπτύχθηκε και εκτελέστηκε μία βιοπληροφορική ανάλυση πειραμάτων με CRISPR αποσιώπηση, η οποία αξιοποιεί το εργαλείο MAGeCK και προσανατολίζεται στα miRNAs.

Σε τεχνικό επίπεδο, με τη σύγκριση των ξεχωριστών αναλύσεων MAGeCK μόνο με miRNA εγγραφές και των τυπικών αναλύσεων MAGeCK (με γονιδιακές και miRNA εγγραφές), συμπεράναμε ότι δεν προκύπτουν σημαντικές αποκλίσεις. Έτσι, καθίσταται εφικτή και λειτουργική η διεξαγωγή ξεχωριστών αναλύσεων για τα miRNAs καθώς και για τους στόχους τους με σκοπό να διασώζονται από την διαδικασία FDR περισσότερα miRNAs/γονίδια. Αναδείχτηκε ακόμα ότι οι βιβλιοθήκες που χρησιμοποιήθηκαν (GeCKO v2), έχουν σχεδιαστεί κατάλληλα ώστε να αποσιωπούν ξεχωριστά τα miRNAs και τα αντίστοιχα γονίδια που εδράζονται στον ίδιο γονιδιακό τόπο, καθώς δεν εμφανίζεται τάση παρόμοιας επιλογής μεταξύ intragenic miRNAs και host γονιδίων.

Αναφορικά με τη λειτουργική σχέση που φέρουν τα miRNAs έναντι των γονιδίων-στόχων τους, χρησιμοποιώντας το σύνολο των γνωστών προβλέψεων στόχων των miRNAs (DIANA-microT-CDS), αναδείξαμε πως δε διαφαίνεται σημαντική τάση των αρνητικά επιλεγμένων miRNAs να ενισχύουν περισσότερο θετικά επιλεγμένους στόχους και αντίστροφα. Ωστόσο, ελέγχοντας αποκλειστικά τους σημαντικά επιλεγμένους στόχους και τα σημαντικά επιλεγμένα miRNAs, αναδεικνύονται μεμονωμένες περιπτώσεις ζευγών miRNA-γονιδίου με πιθανώς σημαντικό ρόλο στον υπό μελέτη φαινότυπο. Μηχανιστικές μελέτες τέτοιων ζευγών ίσως αναδείξουν miRNAs και γονίδια που θα μπορούσαν να αποτελέσουν θεραπευτικούς στόχους και/ή βιοδείκτες που να προβλέπουν απόκριση σε θεραπείες. Τέλος, αναδεικνύοντας μονοπάτια και μοριακές διεργασίες που εμπλουτίζονται από κοινού από επιλεγμένα miRNAs και γονίδια, είναι δυνατόν να διατυπωθούν πιθανές ερμηνείες και βιολογικά ερωτήματα για μελλοντική ερμηνεία των μηχανισμών.

Η παρούσα μελέτη αποτελεί την πρώτη ανάλυση κατά την οποία **(α)** αξιοποιούνται, συνδυαστικά, αποτελέσματα για σημαντικά επιλεγμένα miRNAs και γονίδια από δοκιμασίες διαλογής CRISPR αποσιώπησης, **(β)** πραγματοποιείται έλεγχος των χαρακτηριστικών των sgRNAs που στοχεύουν miRNAs και **(γ)** προτείνονται βιοπληροφορικές ροές για την ανάδειξη miRNAs και στόχων τους με κρίσιμους ρόλους σε μελετούμενους φαινότυπους.

Αναφορές

1. Doudna, J.A. and E. Charpentier, *The new frontier of genome engineering with CRISPR-Cas9*. Science, 2014. **346**(6213): p. 1258096.
2. Jiang, F. and J.A. Doudna, *CRISPR-Cas9 structures and mechanisms*. Annu Rev Biophys, 2017. **46**(1): p. 505-529.
3. Jiang, F. and J.A. Doudna, *The structural biology of CRISPR-Cas systems*. Current opinion in structural biology, 2015. **30**: p. 100-111.
4. Ran, F., et al., *Genome engineering using the CRISPR-Cas9 system*. Nature protocols, 2013. **8**(11): p. 2281-2308.
5. Lin, S., et al., *Enhanced homology-directed human genome engineering by controlled timing of CRISPR/Cas9 delivery*. elife, 2014. **3**: p. e04766.
6. Zhang, H., et al., *Application of the CRISPR/Cas9-based gene editing technique in basic research, diagnosis, and therapy of cancer*. Molecular Cancer, 2021. **20**(1): p. 1-22.
7. Jason, S. and K. Yusa, *Genome-wide CRISPR-Cas9 screening in mammalian cells*. Methods, 2019. **164**: p. 29-35.
8. Li, W., et al., *MAGECK enables robust identification of essential genes from genome-scale CRISPR/Cas9 knockout screens*. Genome biology, 2014. **15**(12): p. 1-12.
9. Wang, B., et al., *Integrative analysis of pooled CRISPR genetic screens using MAGECKFlute*. Nature protocols, 2019. **14**(3): p. 756-780.
10. Jiang, P., et al., *Network analysis of gene essentiality in functional genomics experiments*. Genome biology, 2015. **16**(1): p. 1-10.
11. Zhang, J., *Positive selection, not negative selection, in the pseudogenization of rcsA in Yersinia pestis*. Proceedings of the National Academy of Sciences, 2008. **105**(42): p. E69-E69.
12. Bartel, D.P., *MicroRNAs: genomics, biogenesis, mechanism, and function*. cell, 2004. **116**(2): p. 281-297.
13. O'Brien, J., et al., *Overview of microRNA biogenesis, mechanisms of actions, and circulation*. Frontiers in endocrinology, 2018. **9**: p. 402.
14. Hammond, S.M., *An overview of microRNAs*. Advanced drug delivery reviews, 2015. **87**: p. 3-14.
15. Bushati, N. and S.M. Cohen, *microRNA functions*. Annu. Rev. Cell Dev. Biol., 2007. **23**: p. 175-205.
16. Annese, T., et al., *microRNAs biogenesis, functions and role in tumor angiogenesis*. Frontiers in Oncology, 2020. **10**: p. 581007.
17. Pasquinelli, A.E., S. Hunter, and J. Bracht, *MicroRNAs: a developing story*. Current opinion in genetics & development, 2005. **15**(2): p. 200-205.
18. McDermott, A.M., et al., *The therapeutic potential of microRNAs: disease modulators and drug targets*. Pharmaceutical research, 2011. **28**(12): p. 3016-3029.
19. Slaby, O., et al., *MicroRNAs in colorectal cancer: translation of molecular biology into clinical application*. Molecular cancer, 2009. **8**(1): p. 1-13.
20. Bhaskaran, M. and M. Mohan, *MicroRNAs: history, biogenesis, and their evolving role in animal development and disease*. Veterinary pathology, 2014. **51**(4): p. 759-774.
21. Lewis, B.P., C.B. Burge, and D.P. Bartel, *Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets*. cell, 2005. **120**(1): p. 15-20.
22. Liu, B., et al., *Interplay between miRNAs and host genes and their role in cancer*. Briefings in functional genomics, 2019. **18**(4): p. 255-266.
23. Hinske, L.C.G., et al., *A potential role for intragenic miRNAs on their hosts' interactome*. BMC genomics, 2010. **11**(1): p. 1-13.

24. Shaker, F., et al., *Web-based tools for miRNA studies analysis*. Computers in biology and medicine, 2020. **127**: p. 104060.
25. Kozomara, A., M. Birgaoanu, and S. Griffiths-Jones, *miRBase: from microRNA sequences to function*. Nucleic acids research, 2019. **47**(D1): p. D155-D162.
26. Karagkouni, D., et al., *DIANA-TarBase v8: a decade-long collection of experimentally supported miRNA–gene interactions*. Nucleic acids research, 2018. **46**(D1): p. D239-D245.
27. Reczko, M., et al., *Functional microRNA targets in protein coding sequences*. Bioinformatics, 2012. **28**(6): p. 771-776.
28. Kanehisa, M. and S. Goto, *KEGG: kyoto encyclopedia of genes and genomes*. Nucleic acids research, 2000. **28**(1): p. 27-30.
29. Consortium, G.O., *The Gene Ontology (GO) database and informatics resource*. Nucleic acids research, 2004. **32**(suppl_1): p. D258-D261.
30. Jassal, B., et al., *The reactome pathway knowledgebase*. Nucleic acids research, 2020. **48**(D1): p. D498-D503.
31. Liberzon, A., et al., *Molecular signatures database (MSigDB) 3.0*. Bioinformatics, 2011. **27**(12): p. 1739-1740.
32. Vlachos, I.S., et al., *DIANA-miRPath v3. 0: deciphering microRNA function with experimental support*. Nucleic acids research, 2015. **43**(W1): p. W460-W466.
33. Gibbons, R., et al., *X-linked alpha-thalassemia/mental retardation (ATR-X) syndrome: localization to Xq12-q21. 31 by X inactivation and linkage analysis*. American journal of human genetics, 1992. **51**(5): p. 1136.
34. Wong, L.H., et al., *ATRX interacts with H3. 3 in maintaining telomere structural integrity in pluripotent embryonic stem cells*. Genome research, 2010. **20**(3): p. 351-360.
35. Cavalcante, S.G., et al., *The chromatin remodeler complex ATRX-DAXX-H3. 3 and telomere length in meningiomas*. Clinical Neurology and Neurosurgery, 2021. **210**: p. 106962.
36. Haase, S., et al., *Mutant ATRX: uncovering a new therapeutic target for glioma*. Expert opinion on therapeutic targets, 2018. **22**(7): p. 599-613.
37. Junbo, L., et al., *Genome-wide CRISPR-Cas9 screen reveals selective vulnerability of ATRX-mutant cancers to WEE1 inhibition*. 2019.
38. Chen, A., *PARP inhibitors: its role in treatment of cancer*. Chinese journal of cancer, 2011. **30**(7): p. 463.
39. Jubin, T., et al., *The PARP family: insights into functional aspects of poly (ADP-ribose) polymerase-1 in cell growth and survival*. Cell proliferation, 2016. **49**(4): p. 421-437.
40. Nilov, D., et al., *Modeling of the Enzyme—Substrate Complexes of Human Poly (ADP-Ribose) Polymerase 1*. Biochemistry (Moscow), 2020. **85**(1): p. 99-107.
41. Javle, M. and N. Curtin, *The role of PARP in DNA repair and its therapeutic exploitation*. British journal of cancer, 2011. **105**(8): p. 1114-1122.
42. Juhász, S., et al., *The chromatin remodeler ALC1 underlies resistance to PARP inhibitor treatment*. Science Advances, 2020. **6**(51): p. eabb8626.
43. Tutt, A., et al., *Oral poly (ADP-ribose) polymerase inhibitor olaparib in patients with BRCA1 or BRCA2 mutations and advanced breast cancer: a proof-of-concept trial*. The Lancet, 2010. **376**(9737): p. 235-244.
44. Stern, D.F., *Tyrosine kinase signalling in breast cancer: ErbB family receptor tyrosine kinases*. Breast Cancer Research, 2000. **2**(3): p. 1-8.
45. Hong, L., et al., *Prognostic value of epidermal growth factor receptor in patients with gastric cancer: a meta-analysis*. Gene, 2013. **529**(1): p. 69-72.
46. Voigtlaender, M., T. Schneider-Merck, and M. Trepel, *Lapatinib*. Small molecules in oncology, 2018: p. 19-44.
47. Leinonen, R., et al., *The sequence read archive*. Nucleic acids research, 2010. **39**(suppl_1): p. D19-D21.

48. Rahbari, R., et al., *A novel L1 retrotransposon marker for HeLa cell line identification*. Biotechniques, 2009. **46**(4): p. 277-284.
49. Ning, G., et al., *A novel treatment strategy for lapatinib resistance in a subset of HER2-amplified gastric cancer*. BMC cancer, 2021. **21**(1): p. 1-17.
50. Zhang, L., et al., *Comprehensive Analysis of DNA 5-Methylcytosine and N6-Adenine Methylation by Nanopore Sequencing in Hepatocellular Carcinoma*. Frontiers in cell and developmental biology, 2022. **10**.
51. Zaragoza-Huesca, D., et al., *Identification of Thrombosis-Related Genes in Patients with Advanced Gastric Cancer: Data from AGAMENON-SEOM Registry*. Biomedicines, 2022. **10**(1): p. 148.
52. Chen, F.-Q., et al., *Modification of STIM2 by m6A RNA methylation inhibits metastasis of cholangiocarcinoma*. Annals of Translational Medicine, 2022. **10**(2).
53. Liu, C., et al., *A Functional Reference Map of RNF8 Interactome in Cancers*. 2022.
54. Rao, Z., et al., *A novel HCC prognosis predictor PDSS1 affects the cell cycle through the STAT3 signaling pathway in HCC*. Frontiers in Oncology, 2022. **12**.
55. Kaiyuan, D., et al., *The role and underlying mechanism of miR-1299 in cancer*. Future science OA, 2021. **7**(6): p. FSO693.
56. Lu, J., et al., *The expression and function of miR-622 in a variety of tumors*. Biomedicine & Pharmacotherapy, 2022. **146**: p. 112544.
57. Li, T., X. Sun, and K. Xu, *The suppressing role of miR-622 in renal cell carcinoma progression by down-regulation of CCL18/MAPK signal pathway*. Cell & Bioscience, 2018. **8**(1): p. 1-8.
58. Núñez-Olvera, S.I., et al., *A novel protective role for microRNA-3135b in Golgi apparatus fragmentation induced by chemotherapy via GOLPH3/AKT1/mTOR axis in colorectal cancer cells*. Scientific reports, 2020. **10**(1): p. 1-14.
59. Koleckova, M., et al., *Epithelial to mesenchymal transition and microRNA expression are associated with spindle and apocrine cell morphology in triple-negative breast cancer*. Scientific reports, 2021. **11**(1): p. 1-13.
60. Wang, Y., et al., *miR-3929 Inhibits Proliferation and Promotes Apoptosis by Downregulating Cripto-1 Expression in Cervical Cancer Cells*. Cytogenetic and Genome Research, 2021. **161**(8-9): p. 425-436.
61. Zheng, Y., et al., *Identification of extracellular vesicles-transported miRNAs in Erlotinib-resistant head and neck squamous cell carcinoma*. Journal of cell communication and signaling, 2020. **14**(4): p. 389-402.
62. Qiu, G., et al., *Dysregulation of MALAT1 and miR-619-5p as a prognostic indicator in advanced colorectal carcinoma*. Oncology letters, 2016. **12**(6): p. 5036-5042.
63. Kim, D.H., et al., *Tumor-derived exosomal miR-619-5p promotes tumor angiogenesis and metastasis through the inhibition of RCAN1*. 4. Cancer letters, 2020. **475**: p. 2-13.
64. Xu, W.-T., Z. Yang, and N.-H. Lu, *Roles of PTEN (Phosphatase and Tensin Homolog) in gastric cancer development and progression*. Asian Pacific Journal of Cancer Prevention, 2014. **15**(1): p. 17-24.
65. Wang, S., et al., *miR-107 regulates tumor progression by targeting NF1 in gastric cancer*. Scientific reports, 2016. **6**(1): p. 1-8.
66. Guo, Y. and Y. He, *Comprehensive analysis of the expression of SLC30A family genes and prognosis in human gastric cancer*. Scientific reports, 2020. **10**(1): p. 1-22.
67. de Souza, C.R.T., et al., *MYC deregulation in gastric cancer and its clinicopathological implications*. PloS one, 2013. **8**(5): p. e64420.
68. Wei, C.-Y., et al., *Expression of CDKN1A/p21 and TGFBR2 in breast cancer and their prognostic significance*. International journal of clinical and experimental pathology, 2015. **8**(11): p. 14619.

69. Huang, Y., et al., *Construction of an 11-microRNA-based signature and a prognostic nomogram to predict the overall survival of head and neck squamous cell carcinoma patients*. BMC genomics, 2020. **21**(1): p. 1-11.
70. Li, J., et al., *MiR-186-5p upregulation inhibits proliferation, metastasis and epithelial-to-mesenchymal transition of colorectal cancer cell by targeting ZEB1*. Archives of biochemistry and biophysics, 2018. **640**: p. 53-60.
71. Shi, F., et al., *Retracted Article: RNA-sequencing identified miR-3681 as a negative regulator in the proliferation and migration of cervical cancer cells via the posttranscriptional suppression of HGFR*. RSC advances, 2019. **9**(39): p. 22376-22383.
72. Sun, H., et al., *LncRNA CRNDE promotes cell proliferation, invasion and migration by competitively binding miR-384 in papillary thyroid cancer*. Oncotarget, 2017. **8**(66): p. 110552.
73. Wang, G., et al., *Long non-coding RNA CRNDE sponges miR-384 to promote proliferation and metastasis of pancreatic cancer cells through upregulating IRS 1*. Cell proliferation, 2017. **50**(6): p. e12389.
74. Fan, N., et al., *MicroRNA-384 represses the growth and invasion of non-small-cell lung cancer by targeting astrocyte elevated gene-1/Wnt signaling*. Biomedicine & Pharmacotherapy, 2017. **95**: p. 1331-1337.
75. Yao, Y., et al., *Luteolin suppresses colorectal cancer cell metastasis via regulation of the miR-384/pleiotrophin axis*. Oncology reports, 2019. **42**(1): p. 131-141.
76. Wei, Z., et al., *LncRNA HOTAIR promotes the growth and metastasis of gastric cancer by sponging miR-1277-5p and upregulating COL5A1*. Gastric Cancer, 2020. **23**: p. 1018-1032.
77. Weng, X., et al., *HOTAIR/miR-1277-5p/ZEB1 axis mediates hypoxia-induced oxaliplatin resistance via regulating epithelial-mesenchymal transition in colorectal cancer*. Cell Death Discovery, 2022. **8**(1): p. 310.
78. Zou, W., J.D. Wolchok, and L. Chen, *PD-L1 (B7-H1) and PD-1 pathway blockade for cancer therapy: Mechanisms, response biomarkers, and combinations*. Science translational medicine, 2016. **8**(328): p. 328rv4-328rv4.
79. Zhou, S., et al., *Anti-tumour potential of PD-L1/PD-1 post-translational modifications*. Immunology, 2022.
80. Fan, G., et al., *DNA methylation biomarkers for hepatocellular carcinoma*. Cancer cell international, 2018. **18**(1): p. 1-13.
81. Yang, Z., et al., *Neurotrophin3 promotes hepatocellular carcinoma apoptosis through the JNK and P38 MAPK pathways*. International Journal of Biological Sciences, 2022. **18**(15): p. 5963-5977.
82. Hartwell, H.J., et al., *Prolactin prevents hepatocellular carcinoma by restricting innate immune activation of c-Myc in mice*. Proceedings of the National Academy of Sciences, 2014. **111**(31): p. 11455-11460.
83. Liu, X. and K.-M. Chu, *E-cadherin and gastric cancer: cause, consequence, and applications*. BioMed research international, 2014. **2014**.